Discovering Signals in Noise

Dissertation on the Performance Assessment of Geophysical Instruments and the Contribution of Microgravity Campaigns to Volcano Monitoring

Koymans, M.R.

DOI
10.4233/uuid:65cfc667-d754-41ad-b7c9-6b9a181177e0

Publication date
2024

Document Version
Final published version

Citation (APA)
Koymans, M. R. (2024). *Discovering Signals in Noise: Dissertation on the Performance Assessment of Geophysical Instruments and the Contribution of Microgravity Campaigns to Volcano Monitoring*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:65cfc667-d754-41ad-b7c9-6b9a181177e0

Discovering Signals in Noise

Dissertation on the Performance Assessment of Geophysical Instruments and the Contribution of Microgravity Campaigns to Volcano Monitoring

Discovering Signals in Noise

Dissertation on the Performance Assessment of Geophysical Instruments and the Contribution of Microgravity Campaigns to Volcano Monitoring

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
Thursday, 10 October 2024 at 15:00 o'clock

by

Mathijs Reinier KOYMANS

Master of Science in Earth Sciences, Utrecht University, Utrecht
born in Houten, The Netherlands

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof.dr. L. G. Evers	Delft University of Technology, promotor

Independent members:

Prof.dr.ir. R.F. Hanssen	Delft University of Technology
Prof.dr.ir. C.P.A. Wapenaar	Delft University of Technology
Prof.dr. D.J.J. van Hinsbergen	Utrecht University
Dr.ir. A.A. Verhagen	Delft University of Technology
Dr. M. Poland	United States Geological Survey

Other member:

Dr. E. de Zeeuw-van Dalfsen	Delft University of Technology

TU Delft

Royal Netherlands Meteorological Institute
Ministry of Infrastructure and Water Management

Keywords: Campaign, Microgravity, Data Quality, Performance Assessment, Gravimeters, Seismometers, Geophysical, Volcano, Monitoring Infrastructure

Cover illustration: M. R. Koymans – a Signal Hidden in Noise. Illustration created with speck https://github.com/lucashadfield/speck.

ISBN 978-94-6366-913-9

Typesetting system: LaTeX; Template provided by P. Smets.
Printed in The Netherlands by Gildeprint B.V., Enschede.

MIX
Paper from responsible sources
FSC
www.fsc.org

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Mente et Malleo

Preface

"In the beginning the Universe was created. This has made a lot of people very angry and been widely regarded as a bad move."

Douglas Adams

Somehow back in 2013 I was hired as a student assistant at The Royal Netherlands Meteorological Institute (KNMI), working for ORFEUS Data Center. This opportunity – for which I am still grateful – was probably one of the best introductions a student could have to a professional career in geophysics. After graduating in paleomagnetism in 2015, I continued working at KNMI on the European Plate Observing System, feeling quite satisfied with my first permanent position. But after four years I felt drawn to pursue research once more. At the same time, an interesting position opened up in the NEWTON-g project, designed to study the temporal evolution of Mt. Etna using microgravity observations, a geophysical technique that I did not even know existed in the first place. Within the framework of this project my doctoral research was to be completed and perhaps unsurprisingly, challenges were faced over the duration of the project. In light of this, changes to the scope of the dissertation were warranted. This slight deviation allowed me to interact with many different groups and talented people that I have developed great respect for. The dissertation still mainly concerns the application and improvement of campaign microgravity, but also covers the performance assessment of alternative geophysical instruments commonly used in volcano monitoring infrastructures. Quite evidently I prefer to consider the resulting expansive scope of the dissertation as one of its strengths. Writing the thesis has been challenging, often gratifying, and periodically frustrating – but is something I am proud to present.

Mathijs Koymans,
De Bilt, Sept 18th, 2024

Title Rationale

An effective dissertation title is one that accurately represents the content of the full dissertation, which admittedly, can be tricky to capture in a single coherent sentence. With the apparent need for this rationale in the first place, the effectiveness of the title likely leaves something to be desired. Nevertheless, the title of this dissertation – *Discovering Signals in Noise* – should be considered as being twofold. What are considered signals and what is noise depends on the requirements of the study, and naturally on who the question is presented to. Conventionally, noise is often referred to as that what is undesirable, and after its elimination, what remains can be considered the signal of interest. The first perspective is that even commonly unwanted signals can be leveraged to our benefit, as demonstrated in chapters 2 and 3 – rediscovering noise as a potential signal of interest. In these chapters, the distribution of (ambient) noise in power spectral density estimates from geophysical data, and electrical noise originating from the mains powers grid are used to assess the performance of geophysical instruments commonly operated in volcano monitoring infrastructures. The second perspective may be considered more literally, and is based on the fact that microgravity signals produced by volcanic processes are often minuscule and obscured by ambient and instrumental noise. This presents the challenge of discovering and isolating such volcanic signals in the presence of unavoidable noise and uncertainty, as is extensively encountered in the microgravity studies presented in chapters 4 and 5.

Hence the title of this dissertation – *Discovering Signals in Noise* – either in an attempt to isolate signals from noise, or by acknowledging that noise can in fact be considered a valuable signal too.

Summary

Discovering Signals in Noise - On the Performance Assessment of Geophysical Instruments and the Contribution of Microgravity Campaigns in Volcano Monitoring

Hazards that arise from volcanic settings are abundant, and the continuous monitoring of volcanoes remains an important task to protect civilians. Early warning systems have an enormous impact in effectively informing the general public of potential hazards and reducing the associated risks involved with volcanic eruptions. Volcano monitoring infrastructures rely on various complementary geophysical techniques, and the maintenance of such a diverse system of instruments is often challenging. This dissertation concerns the automated assessment of geophysical instruments, and the application of microgravity observations to long-term volcano monitoring.

Research Framework and Questions

Chapter 1 of this dissertation presents a general introduction that provides an overview of data quality and the fundamentals of (campaign) microgravity analysis and its application to volcano monitoring. After that, the initial research framework and scope of the dissertation are presented. Due to unforeseen obstacles encountered during the project, changes to the desired scope of the dissertation were mandated. One recurring theme in this dissertation is data quality, a motive that is exhaustively encountered in all four main chapters. No matter how evident the desire for the highest data quality appears, it is decisively fundamental to guarantee the validity of results that are used in the evaluation of potential volcanic hazards and associated risks. Microgravity is only one of many techniques available in the multi-disciplinary world of volcano monitoring. For this reason, studies that concern the assessment of various geophysical instruments that are commonly employed in volcano monitoring

infrastructures became an integral part of the dissertation. After the introduction, this dissertation presents two chapters on the assessment of operational geophysical instruments that are commonly employed in volcano monitoring networks. The research question addressed in these two chapters concerns how reliable systems can be developed to automatically assess the quality of geophysical instruments that are commonly deployed in volcano monitoring infrastructures. The following two chapters present studies of long-standing campaign microgravity records from Kīlauea, Hawaiʻi (2009 – 2017) and Askja, Iceland (1988 – 2022) respectively. Microgravity measurements are sensitive to subsurface mass changes induced by hydrothermal or magmatic activity and provide insight into volcanic processes that may otherwise remain concealed. These microgravity chapters address the research questions **(i)** what are the optimal strategies for the collection and treatment of campaign microgravity data, and **(ii)** what additional insights do the results provide in the context of the observed volcanic activity. The added benefit of campaign microgravity in long-term volcano monitoring is demonstrated and discussed, and practical recommendations are provided on how the technique can be applied more reliably.

Per Chapter Summary

Chapter 2 presents a method for the automated analysis of power spectral density (PSD) estimates that is used as an operational product at the Royal Netherlands Meteorological Institute (KNMI) for the assessment of geophysical instruments deployed in the field. Geophysical recordings are often represented by one-dimensional discrete time series that can be decomposed in a collection of harmonics of different frequencies. This representation of the waveform in the frequency domain provides another useful perspective of these data. The distribution of power per frequency provided by PSD estimates is a measure that often remains stable and roughly predictable for an operational instrument through time. For example near coastal regions, PSD estimates of seismometers should contain an expression of ocean-wave interactions (microseism) that generates faintly detectable tremors. Deviations from such anticipated measurements outside of common limits often indicate that the instrument is experiencing degraded performance and not recording such omnipresent signals as expected. In this chapter, multiple theoretically and data-driven constraints on the PSD are presented that can assist with the automated identification of instrumental problems. The standardized method is demonstrated on eight different types of geophysical instruments that are commonly employed in volcano monitoring networks.

Chapter 3 presents a technique for the assessment of geophysical instruments that is demonstrated using data from seismometers and acoustic sensors deployed in the Netherlands Seismic and Acoustic Network (NSAN). The electrical network frequency (ENF) is the frequency at which the alternating current of the electrical grid nominally operates (50 Hz in Continental Europe and 60 Hz in the United States). The ENF may introduce minuscule parasitic signals in geophysical instruments that can be blown up to dominant proportions when amplifiers are involved. While this noise* at the nominal frequency is often easily eliminated using a bandstop filter, this chapter explores the benefits of its presence in geophysical data instead. Random mHz fluctuations from the nominal value are introduced by a perpetual effort of electrical grid operators to balance the electrical supply with the required demand. These variations in grid frequency are consistent over the entire electrical grid and represent a nation – or even continent wide – time calibration signal. The chapter demonstrates that the signal can be extracted from spectrograms of geophysical data and allows for timing corrections to be near the 1 s level. A completely separate application of the ENF allows for orientation anomalies to be detected. This utility is facilitated by an observed highly polarized response to the dominant source that introduces the ENF to the data. Such source – which may be a transformer or nearby electrical power supply – is often stable and does not move with time. Using the technique, an unintentionally misaligned surface accelerometer (rotated by 90°) was identified in the Netherlands Seismic and Acoustic Network (NSAN) and realigned. Further applications of this technique are foreseen, and multiple avenues exist for future research on this interesting, albeit quite unusual topic.

Chapter 4 presents a study of a near decade-long (2009 – 2017) campaign microgravity record from Kīlauea, Hawai'i. This record spans an eruption that featured a receding and advancing lava lake, that terminated in May 2018 with a devastating outflow in a residential neighbourhood in the Lower East Rift Zone (LERZ). This outflow was followed months later by the collapse of Kīlauea's summit crater. The microgravity data are analysed using a weighted least squares inversion, solving for gravity differences, instrumental drift, and a potential gravimeter tare at the same time. The proposed method in this dissertation (appendix A) is implemented in software and suggested as the most effective way to treat campaign microgravity data [*Reilly*, 1970; *Hwang et al.*, 2002; *Hector and Hinderer*, 2016; *Koymans*, 2022a]. It is recommended that microgravity campaigns are completed in a way so that they optimally utilise the inherent advantages of the proposed method. The Kīlauea campaign microgravity record is diligently maintained and key lessons can be taken away from this chapter as to how campaign microgravity data should be collected and treated in a volcanic setting. The high spatial resolution of the Kīlauea microgravity network also facilitates the inversion of microgravity results to point masses, of which the source parameters can be estimated effectively. Campaign microgravity results in combination with continuous microgravity data [*Carbone et al.*, 2013] indicate that the density of magma accumulating in the Halema'uma'u reservoir may have been steadily increasing, potentially leading up to the destructive outflow in the lower ERZ in 2018.

Chapter 5 concerns a campaign microgravity study of Askja, an Icelandic volcanic complex that suddenly returned to the spotlight in August 2021. The volcano was characterised by an extended period of exponentially decaying subsidence dating back to at least 1984 [*Sturkell et al.*, 2006]. The polarity of deformation suddenly

*In accordance with the supporting title rationale, the term *noise* should probably be replaced with *signal* here.

inverted in August 2021 and the caldera floor began rising at a rate of approximately 40 cm per year, without an indication of slowing down[*]. Askja is characterised by a uniquely long record of campaign microgravity data with campaigns episodically completed since 1988. In this chapter, the most recent microgravity campaigns from 2021 and 2022 were completed after the caldera floor started rising and are considered in a historical context, covering the turnaround from caldera subsidence to uplift. Microgravity data indicate that mass has been continuously moving away from the caldera center between 1988 and 2017. Data from the 2021 campaign show that mass had likely accumulated below the center caldera since 2017, but this accumulation cannot be timed precisely and may have occurred progressively leading up to, or just during the period of uplift. It is therefore impossible to determine whether microgravity results could have served as a precursor to the renewed activity observed at Askja. During the period of uplift between the campaigns of 2021 and 2022, microgravity changes follow the free-air gradient, suggesting that a subsurface density decrease of the previously accumulated mass may drive the persistent uplift. This study also emphasizes the benefit of completing campaigns at regular intervals and in consistent fashion so that derived results can be used more reliably in long-term volcano monitoring.

Chapter 6 concerns the synthesis and reconciles the preceding chapters, summarizes the findings and shortcomings of the dissertation, and presents practical recommendations on how the collection and treatment of campaign microgravity can be improved in volcano monitoring. Furthermore, a description of the integration of continuous absolute measurements from a quantum gravimeter installed at the summit of Mt. Etna with the European Integrated Data Archive (EIDA) is provided. This archive is a modern and FAIR compliant data infrastructure that facilitates automated access to multi-disciplinary geophysical data – a feature that is extremely desirable and fundamental for early warning volcano monitoring systems.

[*] It slowed down briefly in September 2023 and uplift continued in November 2023.

Samenvatting

Het ontdekken van signalen in ruis: proefschrift over de evaluatie van het prestatievermogen van geofysische instrumenten en de bijdrage van microzwaartekrachtmetingen aan vulkaanwaarnemingen.

Gevaren die voortkomen uit vulkanische activiteit zijn veelvuldig, en het voortdurend waarnemen van vulkanen blijft een belangrijke taak in het waarborgen van de volksveiligheid. Waarschuwingssystemen hebben een belangrijke rol in het effectief en tijdig inlichten van de maatschappij over mogelijke gevaren, om zodoende de gevolgen van grotendeels voorkombare risicos zo klein mogelijk te maken. Waarneeminfrastructuur berust vaak op informatie van veel verschillende meetapparatuur, en het bijhouden en verifiëren van de betrouwbaarheid van deze meetapparatuur en informatie is veeleisend. Dit proefschrift behandelt de geautomatiseerde beoordeling van geofysische meetapparatuur in het veld, en de bijdrage van microzwaartekrachtmetingen op vulkaanwaarnemingen.

Onderzoekskader en Onderzoeksvragen

Hoofdstuk 1 introduceert het proefschrift en geeft een overzicht van de datakwaliteit van geofysische meetapparatuur en behandelt de basisprincipes van microzwaartekrachtmetingen en hun bijdrage aan vulkaanwaarnemingen. Dit wordt vervolgd met de opzet van het onderzoek, inclusief de onderzoeksvragen en het onderzoekskader van het proefschrift. Door onvoorziene omstandigheden is de afbakening van dit proefschrift tijdens het onderzoek enigzins verlegd en aangepast. Datakwaliteit bleek een terugkerend thema te zijn in dit proefschrift en staat in ieder hoofdstuk centraal. Hoge datakwaliteit wenselijk is omdat het een fundamenteel onderdeel voor een goede inschatting van vulkanische gevaren en geassocieerde risico's bevorderd. Daarnaast is microzwaartekracht maar een van de vele beschikbare methoden om vulkaanwaarnemingen te voltooien. Dit gaf aanleiding twee hoofdstukken aan het proefschrift toe te voegen betreffende het onderzoek naar datakwaliteit van verschillende geofysische meetinstrumenten. Aansluitend op de introductie bevat dit proefschrift dus twee hoofdstukken die betrekking hebben op het vaststellen van het prestatievermogen van geofysische meetapparatuur die veelvuldig worden gebruikt tijdens vulkaanwaarnemingen. De daaropvolgende twee hoofdstukken hebben betrekking op de analyse van langlopende microzwaartekracht registraties die respectievelijk zijn genomen op Kīlauea, Hawai'i (2009 - 2017) en Askja, IJsland (1988 - 2022). Microzwaartekrachtmetingen zijn gevoelig voor ondergrondse massaveranderingen, welke worden veroorzaakt door hydrothermale of magmatische activiteit en en leveren op die manier een bijdrage aan het waarnemen van vulkanische processen die anders mogelijk niet konden worden beschouwd. Deze hoofdstukken tonen de essentiele bijdrage van zwaartekrachtscampagnes aan voor het langdurig waarnemen van vulkanische activiteit. De laatste hoofdstukken bevatten praktische aanbevelingen voor hoe deze technieken effectief en betrouwbaar kunnen worden toegepast voor dit doeleinde.

Samenvatting per Hoofdstuk

Hoofdstuk 2 beschrijft een toepasbare methode voor de geautomatiseerde beoordeling van *power spectral density* (PSD) schattingen. Deze methode wordt momenteel gebruikt als een operationeel product binnen het Koninklijk Nederlands Meteorologisch Instituut (KNMI) om de betrouwbaarheid van instrumenten in het veld vast te stellen. Geofysische metingen betreffen vaak eendimensionale discrete tijreeksen die kunnen worden ontbonden in een verzameling van harmonische trillingen van verschillende frequenties. Deze beschouwing van de informatie in het frequentiedomein biedt een nuttig alternatief om de datakwaliteit van geofysische metingen te beoordelen. Namelijk, de verdeling van het vermogen over de frequenties in het gemeten signaal wordt gegeven door de PSD, en blijft relatief stabiel wanneer het meetinstrument operationeel is. Bijvoorbeeld, nabij de kust wordt verwacht dat de atmosfeer en oceaan geringe trillingen veroorzaken, genaamd microseism, die meetbaar zijn voor het instrument. Enige afwijking van zulk alom aanwezig en verwacht achtergrondruis kan aangeven dat het instrument niet optimaal functioneert. In dit hoofdstuk worden verschillende drempelwaarden aan de PSD opgelegd die niet – of juist wel – mogen worden overschreden. Deze gestandaardiseerde methode wordt toegepast op acht verschillende soorten geofysische meetapparatuur die veelvuldig worden gebruikt voor vulkaanwaarnemingen.

Hoofdstuk 3 presenteert een techniek om datakwaliteit te beoordelen afkomstig van apparatuur die wordt ingezet in het Nederlandse Seismische en Akoestische Netwerk (NSAN). De stroomnetfrequentie (ENF) van de wisselstroom die wordt gebruikt op het stroomnet is nominaal gesproken stabiel (50 Hz binnen Europa en 60 Hz binnen de Verenigde Staten). De ENF kan geringe lekstromen in meetapparatuur veroorzaken die door interne versterkers worden uitvergroot tot significante amplitudes. Ondanks dat deze ruis* rondom de nominale frequentie makkelijk te verwijderen is met een bandstop filter, beschouwt dit hoofdstuk de voordelen van

*In overeenstemming met de titel van dit proefschrift zou hier de term signaal moeten worden gebruikt.

de aanwezigheid van het signaal. Willekeurige mHz fluctuaties rondom de nominale frequentie worden geïntroduceerd door netwerkbeheerders die een precieze balans proberen te vinden tussen stroomaanbod en stroomverbruik. Deze variaties in de stroomnetfrequentie zijn gelijktijdig meetbaar binnen het gehele stroomnet en bieden dus een landelijk, of zelfs continentaal, calibratiesignaal. Dit hoofdstuk demonstreert dat deze kleine veranderingen kunnen worden onttrokken uit geofysische metingen, en dat zodoende tijdanomalien op het niveau van 1 s kunnen worden opgespoord. Een bijkomend voordeel is dat mogelijke oriëntatiefouten van meetapparatuur in het veld kunnen worden bepaald. Dit is mogelijk door de hoge mate van lineariteit in het signaal die consistent is met een nabijgelegen bron. Van deze bron, welke mogelijk een transformator of stroomvoorziening is, wordt aangenomen dat deze door de tijd niet beweegt. Zodoende werd met deze methode een onjuist geïnstalleerde accelerometer (geroteerd met 90°) ontdenkt in het NSAN en gecorrigeerd. Er worden meedere mogelijke toepassingen voorzien die gebruikmaken van deze interessante doch ongebruikelijke toepassing.

Hoofdstuk 4 betreft een onderzoek naar decennialange activiteit van Kīlauea, Hawaiʻi, waarvoor een archief van microzwartekrachtdata beschikbaar is (2009 – 2017). Dit archief doorloopt de langdurige activiteit welke werd gekenmerkt door een uniek fluctuerend lavameer. De vulkanische activiteit werd uiteindelijk beëindigd met een verwoestend uitbarsting in de Lower East Rift Zone (LERZ). Deze uitbarsting werd maanden later zelfs gevolgd door de instorting van de krater van Kīlauea. De microzwaartekrachtdata zijn geanalyseerd door gebruik te maken een van een keinste-kwadratenmethode, waarbij gelijktijdig een oplossing wordt gezocht voor zwaartekrachtsverschillen, instrumentele drift, en mogelijke gravimeter *tares* (afwijkende sprongen). De voorgestelde methode (appendix A) is geïmplementeerd in beschikbare software en wordt aangeraden als de meest gangbare en effectieve manier om zwaartekrachtsmetingen uit campagnes te verwerken [*Reilly*, 1970; *Hwang et al.*, 2002; *Hector and Hinderer*, 2016; *Koymans*, 2022a]. Daarnaast wordt ook voorgesteld om microzwaartekrachtcampagnes te voltooien op een manier zodat de inherente voordelen van deze methode effectief kunnen worden ingezet. De metingen genomen op Kīlauea zijn aandachtig bewaard, en er kunnen hieruit verschillende lessen worden getrokken over hoe zwaartekrachtsdata gemeten en behandeld dienen te worden. De hoge ruimtelijke resolutie van het Kīlauea zwaartekrachtsnetwerk maakt het mogelijk om de metingen om te zetten naar puntbronnen onder de vulkaan, van welke de broneigenschappen kunnen worden geschat. Dit geeft zodoende een beeld van massaveranderingen onder de vulkaan. De combinatie van campagnemetingen samen met continumetingen [*Carbone et al.*, 2013] geven aan dat tijdens de vulkanische activiteit de dichtheid van het magma door de tijd toenam, en dat deze verzwaring mogelijk heeft bijgedragen aan de vernietigende uitbarsting in de LERZ in 2018.

Hoofdstuk 5 betreft een onderzoek wat gebruik maakt van campagnemetingen afkomstig uit Askja, Iceland, een vulkanisch complex welke in augustus 2021 plotseling terug in de schijnwerpers kwam te staan. De vulkaan werd sinds 1984 gekarakteriseerd door een langdurige periode van exponentieel afnemende bodemdaling. In augustus 2021 stopte de bodemdaling plotseling en begon de vulkaan verticaal uit te zetten met een snelheid van ongeveer 40 cm per jaar, tot zover zonder enig teken van afremming*. Askja wordt gekenmerkt door een uniek archief van zwaartekrachtsmetingen die teruggaan tot 1988 en sindsdien episodisch werden waargenomen, zowel in de laatste jaren. In dit hoofdstuk worden twee nieuwe datasets toegevoegd aan dit bestaande archief welke zijn waargenomen tijdens de periode van bodemstijging. De zwaartekrachtdata worden beschouwd in historische context inclusief het inflectiepunt van bodemdaling naar bodemstijging. De microzwaartekrachtdata laten zien dat massa zich onder de vulkaan heeft terugbewogen tussen 1988 en 2017. De metingen vanaf 2021 laten zien dat opnieuw een significante hoeveelheid massa is teruggeplaatst sinds 2017, maar het is niet mogelijk om in te schatten of dit moment van terugkeer voor of tijdens de periode van bodemstijging plaatsvond. Door ontbrekende metingen rondom het inflectiepunt is het dus niet mogelijk om te bepalen of de zwaartekrachtdata een indicatie hadden kunnen geven van massatoename voordat de bodemstijging werd ingezet. Tijdens de periode van bodemstijging geven de metingen aan dat de stijging plaatstvindt onder een *free-air gradient* regime, wat mogelijk verklaard kan worden door een dichtheidsverlaging onder de vulkaan. Dit onderzoek demonstreert dat het belangrijk is om zwaartekrachtsmetingen regelmatig uit te voeren zodat resultaten betrouwbaarder kunnen worden toegepast voor het langdurig waarnemen van vulkanen.

Het afsluitende hoofdstuk 6 biedt een synthese en conclusie van de voorgaande hoofdstukken en beschrijft ook de tekortkomingen van dit proefschrift. In dit hoofdstuk worden er aanbevelingen gedaan om campagnemetingen zo betrouwbaar mogelijk te laten verlopen. Tevens wordt een beschrijving geboden van de samenvoeging van continumetingen afkomstig van een absolute quantum gravimeter die is geïnstalleerd op de vulkaan Etna binnen het *European Integrated Data Archive* (EIDA). Dit archief is modern, volgt FAIR data management principes, en biedt data aan van multidisplinaire geowetenschappen. Deze functie is zeer waardevol voor vulkaanwaarnemingen en voor het vroegtijdig inschatten en waarschuwen voor mogelijke risicos en gevaren.

*In ieder geval tot juli 2023.

Contents

1

Introduction

1.1 Quality Assessment of Geophysical Instruments and Data

Applied (geophysical) research often begins with the collection of data from the field. Evidently, this is a crucial stage as many scientific results are a direct product of these data and integrate all potential errors made during data collection and treatment. Data that are used in research are also often collected by external parties, as the availability of FAIR repositories have become more commonplace. Researchers may often rely on these data, and assume that they are reliable and well curated, and that accompanying metadata are accurate – perhaps even with blind faith.

Data providers, including researchers publishing data sets in repositories, should feel intrinsically responsible for the quality of the data that they provide. However, this can only be fairly considered to be following a best-effort approach as the collection of data requires diligent and hard work. Ultimately, providers are not exclusively responsible and users should always apply due diligence when utilizing data and software. Science should be self-correcting as errors are unavoidable and even ubiquitous; trust but verify. Not from a perspective of inherent distrust but from that of collaboration.

In light of this, a logical step is the development of tools for the identification and rectification of existing errors in (meta)data. In seismology, various quality control methods and tools exist e.g., PQLX [*McNamara and Boaz*, 2006a], MUSTANG [*Ahern et al.*, 2015], DQA [*Ringler et al.*, 2015], WFCatalog [*Trani et al.*, 2017], AutoStatsQ [*Petersen et al.*, 2019], including component ratio monitoring [*Pedersen et al.*, 2020]. An initial step is calculating quality metrics, and a following step is the automatic detection of anomalies and sending out alerts and performing actual action, whereas the first is more often implemented than the second. Such quality process becomes particularly challenging with an increasing number of operational instruments (e.g., over 850 instruments for the Netherlands Seismic and Acoustic Network [*KNMI*, 1993]). It is not a trivial task for data providers to measure continuously and guarantee high data quality at the same time. Manual inspection of all data has become impractical and with the addition of even greater scale large-N deployments (e.g., *Li et al.* [2018]; *Dougherty et al.* [2019]), the appeal of passive and automated assessment of data has equally become larger too. Validation of data integrity is also critical in case of malicious tampering, for instance during the verification of compliance with the Comprehensive Nuclear-Test-Ban Treaty *Coyne et al.* [2012]. Besides, volcano monitoring infrastructures often integrate a versatile set of data from disciplines including seismology, acoustics, Global Navigation Satellite System (GNSS) observations, tiltmeters, gravimetry, and more. To facilitate effective data curation, it is important to identify shared methods that can be applied to multiples types of instruments. In this dissertation, these problem are considered and tackled using automated analysis of power spectral density (PSD) estimates and through the use of electrical network frequency (ENF) analysis.

1.2 Microgravity

Changes in *gravity* – formally, changes in the vertical component of gravitation acceleration (g) – can be observed with sensitive instruments called gravimeters. The absolute value of acceleration ($g \approx 9.81\,\mathrm{m\,s^{-2}}$ in The Netherlands) is not constant everywhere on Earth. In fact, the difference in gravity between the top of the Dom tower in Utrecht (112 m) and the ground surface can be measured to be approximately 34 500 µGal (where $1\,\mu\mathrm{Gal} = 1 \times 10^{-8}\,\mathrm{m\,s^{-2}}$). The observed decrease in gravity at higher altitudes can be attributed to the increased distance from Earth, following Newton's law of universal gravitation (force being inversely proportional to distance squared). A similar effect can be observed from e.g., the gravitational interaction between the Earth and Moon, inducing detectable periodic changes in gravity. These global effects are easily observed* and integrate the contributions of all mass of the entire Earth (e.g., *Crossley et al.* [2013]). Likewise, any change in mass distribution through time in close proximity to a gravimeter may produce a detectable gravity signal (*Carbone et al.* [2013, 2015]). Because these signals are often of such small amplitude they are referred to as changes in *microgravity*. In the final two chapters of this dissertation, temporal changes in microgravity at volcanoes are studied that may be caused by local displacements of subsurface fluids of hydrothermal or magmatic nature. These observations contribute to the understanding of the subsurface processes that drive volcanic eruptions.

1.3 Measurements of Microgravity

Conventional gravimeters are most sensitive to variations in the distribution of mass vertically above and below the gravimeter. In terrain microgravity, a distinction can be made between two main types of observations:

[1] **Absolute gravimeters** make measurements of the local proper gravitational

*By sensitive gravimeters, at least – but technically also by individuals on a beach casually watching the tide roll in.

acceleration. These instruments measure gravity by dropping either a physical projectile (FG5; *Niebauer et al.* [1995]) – or for the latest generation of instruments – a cloud of supercooled atoms (Absolute Quantum Gravimeters; *Ménoret et al.* [2018]). The trajectory is measured using (atom) interferometry and precisely timed, thereby developing a sensitivity to gravitational acceleration. In quiet environments, these instruments can reach a precision of up to 0.1 µGal ($1\,\mathrm{nm\,s^{-2}}$). Absolute gravimeters often require extensive infrastructure to operate effectively, and are extremely suitable for providing stable and precise absolute gravity reference points.

[2] **Relative gravimeters** measure a difference in gravity from one point to another, or at the same point over time (or a combination thereof). These measurements are thus only meaningful when expressed relative to another dedicated measurement. Most relative gravimeters are based on a mass-on-spring principle, where a change in force exerted on the proof-mass causes a detectable displacement. The displacement at zero frequency (direct component) is directly proportional to the inertial acceleration experienced by the proof-mass [*Havskov and Alguacil*, 2016] relative to its housing, and hence also sensitive to changes in gravity. One key challenge in the development of relative gravimeters are its sensitivity and stability over time (drift). Instruments that are effectively drift free, such as the *iGrav* superconducting gravimeter [*Warburton et al.*, 2010], levitate a niobium sphere in a stable magnetic field but are not portable and require a lot of supporting infrastructure. Campaign microgravity measurements using portable spring-based, battery operated instruments (CG-5, CG-6; *Scintrex Limited* [2012]; *Francis* [2021]) are also not meaningfully affected by instrumental drift, in the sense that the drift can be estimated over a short duration and eliminated. The observed precision of superconducting gravimeter can match that of absolute gravimeters, while the best mechanical spring gravimeters are often limited to 5 µGal [*Scintrex Limited*, 2012].

While continuous absolute gravity measurements generally have a higher accuracy and precision compared to relative measurements, they require significant effort to set up and infrastructure to support them. Relative campaign measurements provide a suitable alternative, allowing for measurements under harsh field conditions over a large spatial region. Because the instruments need to be transported, campaigns require a large time investment and are often completed at yearly intervals at best. Campaign microgravity is thus inherently characterised by a low temporal resolution and particularly appropriate for long-term (multi-year) volcano monitoring.

■ 1.3.1 Discovering Microgravity Signals in Noise

The property that makes gravimeters useful in the first place – their unrivaled sensitivity – also makes them susceptible to many sources of undesirable noise. Omnipresent gravity signals of known period and amplitude (that can thus also be used to assess gravimeter performance) include the effects of the (solid) Earth tides that are caused by the gravitational interaction between the Sun, Moon, Earth, including other planetary bodies. These effects can be accurately modeled (e.g., *Longman* [1959]; *Wenzel* [1996]) and are easily reduced from the observations. Terrestrial effects however, are much harder to estimate, and include inertial (seismic) noise, variations in groundwater level [*Miller et al.*, 2017; *Poland and de Zeeuw-van Dalfsen*, 2019; *Carbone et al.*, 2019], atmospheric pressure [*Van Camp et al.*, 2017; *Francis*, 2021], and to summarize – every other process that involves the redistribution of mass in any way. Isolating volcanic signals from such environmental noise remains a key challenge in the successful application of microgravity observations in volcano monitoring. Because the gravity effect of such environmental sources are difficult to quantify precisely, this dissertation recommends methods which avoid the introduction of significant additional uncertainties from data collection and treatment.

1.4 Application of Microgravity in Volcano Monitoring

Volcanic processes typically involve mass redistribution of subsurface fluids that can be e.g., hydrothermal or magmatic of nature. Figure 1.1 summarizes the wavelengths, timescales, and magnitudes of gravity changes commonly associated with volcanic processes [*Carbone et al.*, 2017; *Nikkhoo et al.*, 2018]. Microgravity observations can therefore assist to provide quantitative estimates of source characteristics at volcanoes worldwide [*Jousset et al.*, 2003; *Carbone and Greco*, 2007; *de Zeeuw-van Dalfsen et al.*, 2013; *Miller et al.*, 2017, 2018; *Poland et al.*, 2019]. Forecasting volcanic events remains one of the most sought-after and challenging pursuits in the study of active volcanoes. It has been suggested in some case that microgravity signals can be interpreted as precursors to eruptive events [*Rymer*, 1994; *Battaglia et al.*, 2008; *Carbone et al.*, 2013], further illustrating the added potential benefit of microgravity observations in volcano monitoring infrastructures.

Observations from Global Navigation Satellite System (GNSS) and Interferometric synthetic-aperture radar (InSAR) provide information on surface deformation. From this information, source volumes and pressures can be estimated (e.g., *Sturkell et al.* [2006]; *Jo et al.* [2015]; *Bemelmans et al.* [2021]). Microgravity data can provide complimentary constraints on the source mass and mechanism responsible for the deformation. In particular, the gradient determined by the change in gravity (Δg) over the change in height (Δh) represents an important parameter [*Rymer*, 1994; *Van Camp et al.*, 2017]. When this gradient approaches the theoretical free-air gradient (FAG) of roughly 308 µGal/m, it indicates that no subsurface mass was added, and the cause of deformation can potentially be attributed to a change in subsurface density. When the gradient approaches the Bouguer corrected free-air gradient (BCFAG), it indicates that no density change was observed but the addition of mass was potentially the source of the uplift. The combination of surface deformation observations (InSAR, GNSS, levelling, and tilt) and terrain microgravity measurements thus compliment each other and provide information on the source of the subsurface changes at volcanoes.

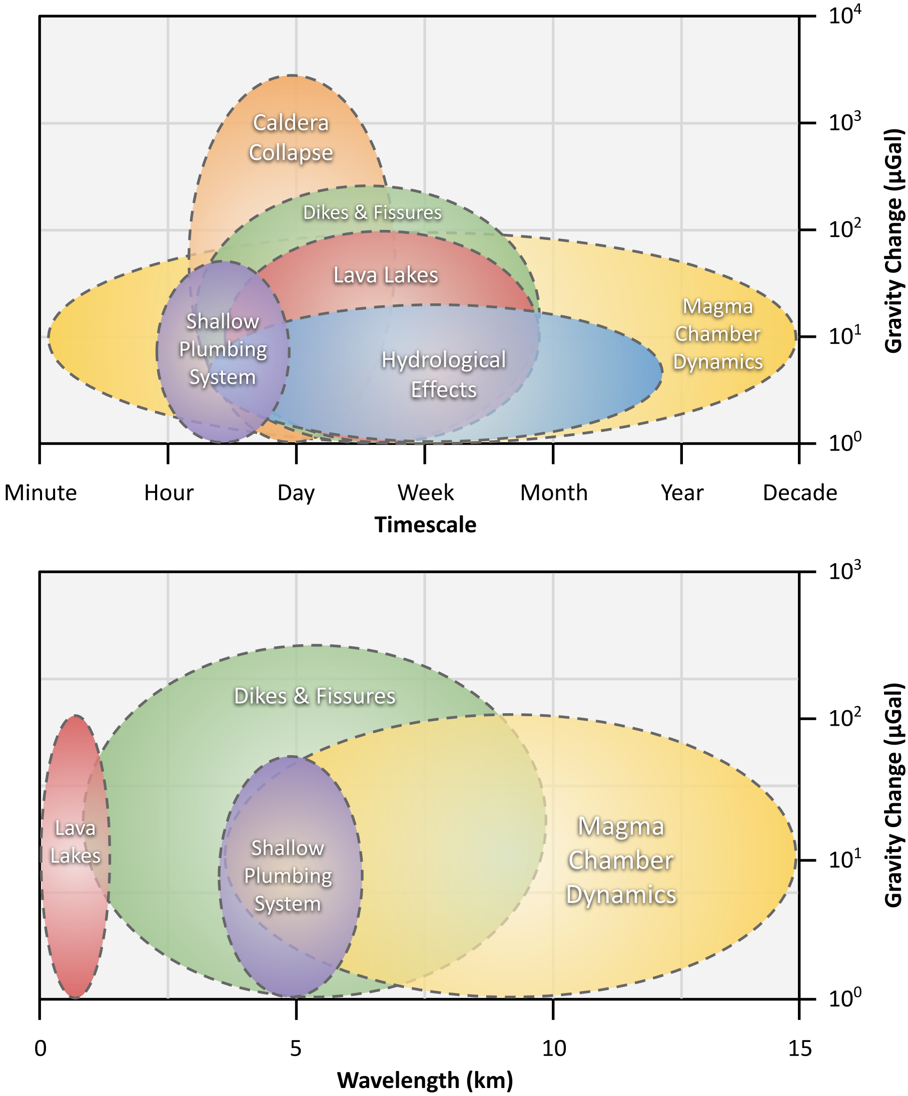

Figure 1.1: *Overview of timescales, wavelengths, and magnitudes of microgravity changes commonly associated with volcanic processes. Figure was modified after Nikkhoo et al. [2018] based on processes described by Carbone et al. [2017] and references therein.*

Microgravity signals commonly associated with processes expected near active volcanoes - fig. 1.1 illustrates timescales, wavelengths, and magnitudes of gravity changes associated with the respective process:

Lava Lakes - Changes in the level of nearby lava lakes may cause rapid mass displacements and high amplitude gravity signals [*Carbone et al.*, 2013].

Caldera Collapse - Rapid disintegration of the caldera floor after magma eviction causes sudden redistribution of material, associated with a gravity decrease [*Furuya et al.*, 2003; *Poland et al.*, 2019].

Shallow Plumbing System - Gas pistoning and lava fountaining may cause a decrease in gravity due to built up gas or a foam layer in the magma chamber [*Carbone et al.*, 2015].

Dikes & Fissures - Intrusions causing deformation in the host-rock and can induce gravity change (e.g., *Okubo* [1991]; *Nikkhoo and Rivalta* [2022]).

Magma Chamber Dynamics - Long-term changes by the accumulation or eviction of magma from subsurface storage areas [*Rymer and Williams-Jones*, 2000].

Hydrological Effects - Hydrothermal and transient seasonal effects (rain, snow), including variations in the level of the water table [*Hinderer et al.*, 2016; *Carbone et al.*, 2019; *Poland and de Zeeuw-van Dalfsen*, 2019].

1.4.1 Current Status of Microgravity Analysis in Volcano Monitoring

At present, the application of microgravity in volcano monitoring is underutilized in an operational setting and only conventionally used at Mt. Etna, Sicily [*Carbone et al.*, 2013] and Kīlauea, Hawaiʻi [*Poland and Carbone*, 2018]. This can be mainly attributed to the limited spatio-temporal resolution that is a direct consequence of high instrumental cost and the effort required to obtain measurements. Using a single gravimeter, there is an inherent trade-off between resolution in time and space. The choice is straightforward: the gravimeter could either be left at one place for an extended period, or moved to various places for a shorter duration. Some of the volcanic processes illustrated in fig. 1.1 require continuous observations over hours (e.g., lava lake dynamics and fountaining) at one location, while other processes happen over a longer period and wavelength and may be more suitable for campaign measurements completed annually (e.g., magma chamber dynamics), illustrating the possible need for a hybrid approach [*Hinderer et al.*, 2016]. It follows that one of the the biggest advances that can be made is to fulfil the desire for continuous measurements at multiple locations. Projects that aim to overcome the challenges in terrain gravimetry attempt to leverage a large number of cost-effective MEMS (Microelectromechanical Systems) gravimeters (e.g., *Middlemiss et al.* [2016]; *Carbone et al.*

[2020]). This approach would employ an order of magnitude more instruments, with the implicit assumption that the benefit of the increased spatio-temporal resolution outweighs the superior performance of a conventional high-cost instrument.

Besides the desired increase in spatio-temporal resolution, another key observation is that campaign microgravity applied to the study of volcanoes are missing standardization in terms of data collection and treatment. There is a need for a comprehensive description on how microgravity campaigns should be completed and treated in an optimal fashion, including mentions of potential pitfalls so that be avoided in future campaigns. An enormous effort can be spent correcting microgravity data for minor source of undesirable effects and interpreting derived results, but obtaining reliable data from the field is the lowest hanging fruit to target first.

1.5 Research Questions, Dissertation Motivation and Outline

This dissertation was conceptualized in the framework of a proposal submitted in the H2020 FET (Future Emerging Technologies) call round. The grant enables projects that aim to innovate, and provides an environment of high risk with high potential reward. Within the funded NEWTON-g project (**New To**ols for Terrai**n G**ravimetry*; *Carbone et al.* [2020]), the proposed goal was to improve the spatio-temporal resolution of microgravity observations on Mt. Etna through the development and deployment of an array of 20 – 30 continuous MEMS gravimeters [*Middlemiss et al.*, 2016]. The individual MEMS gravimeters would represent pixels in a so-called gravity imager – providing insight into volcanic gravity changes in high resolution over space and time, eventually contributing data to real-time volcanic hazard analysis. The MEMS devices were to be anchored through relative campaign measurements to a commercial Absolute Quantum Gravimeter (AQG) in order to eliminate any residual instrumental drift. The AQG was deployed in the project and is presently operational at the Pizzi Deniri volcano observatory, nearby the summit craters of Mt. Etna at 2800 m altitude [*Antoni-Micollier et al.*, 2022]. The main research question to be addressed was to study and quantify the added benefit of an array of continuous gravimeters on an active volcano in terms of characterising volcanic sources (fig. 1.1).

Consequences of the COVID-19 Pandemic

As a direct consequence of lockdowns imposed by national governments in response to the COVID pandemic, the design and assembly of the MEMS devices was severely delayed. On top of that, even though impressive advances have been made in improving the stability of the instruments during the project [*Anastasiou et al.*, 2022], up till now it remains challenging to detect gravimetric changes associated with volcanic processes using MEMS gravimeters. The contribution of instrumental drift remains a limiting factor, and is likely imparted by thermal and pressure variations because the device is not fully insulated from external influences. Because of the unavailability of sufficient reliable MEMS gravimeter data for analysis, the scope of

*At some point catchy project *acronyms* have evidently become a little dubious.

the dissertation was adjusted. To fill this hiatus, campaign microgravity measurements were introduced and studied, and data quality became a central theme of the dissertation.

Adapted Scope of the Dissertation and Research Questions

A positive side effect of the challenges encountered was that the importance of data quality could be highlighted, specifically for microgravity data, but also for alternative geophysical measurement techniques that are used in volcano monitoring. A logically following research question was whether it was possible to develop methods for the passive and automated detection of degraded performance of such instruments. Chapter 2 outlines a study, presenting a technique that describes the development and implementation of an operational product through which the data quality from geophysical instruments can be assessed through an automated analysis of power spectral density (PSD) estimates [*Koymans et al.*, 2021]. The methods that are presented and discussed in chapter 3 on the application of the Electrical Network Frequency (ENF) serve a similar purpose, with a particular focus on the identification of timing issues and instrumental orientation anomalies (**Research Question I**)

> **Research Question I** – How can we develop reliable methods for the passive assessment of geophysical instruments employed in volcano monitoring infrastructures?
>
> Chapters 2 and 3 of this dissertation

Chapter 4 concerns a study of campaign microgravity data at Kīlauea, Hawaiʻi [*Koymans et al.*, 2022], spanning a near decade-long eruption from 2008 – 2018. The focus of chapter 5 is on Askja, an Icelandic volcano that shows rapid uplift since August 2021 at a rate of approximately 40 cm/yr [*Parks et al.*, 2022] after forty years of continuous subsidence. The research questions addressed in these chapters concern whether changes in microgravity can be related to the observed surface deformation, and whether these results can provide additional constraints and insights into the driving subsurface processes (**Research Question II**).

> **Research Question II** – What additional insights do campaign microgravity provide about the 2008 – 2018 eruption of Kīlauea, Hawaiʻi, and the transition from caldera subsidence to uplift in August 2021 at Askja, Iceland?
>
> Chapters 4 and 5 of this dissertation

The two chapters on microgravity also provide fundamental recommendations for the collection and treatment of microgravity data. Despite microgravity campaigns being completed since the late 1980's (e.g., *Rymer and Brown* [1989]; *Rymer*

[1994]), it became apparent there was a need for agreement and standardization in data collection and treatment. It is demonstrated that there exists a large inherent variability in the results, that is caused by different campaign strategies [*Murray and Tracey*, 2001] and the chosen approach for analysis [*Hector and Hinderer*, 2016; *Battaglia et al.*, 2022]. It would be beneficial for the volcano gravimetry community to agree on these strategies, so that uncertainties associated with data collection and treatment can be largely eliminated. Microgravity observations are inherently uncertain and their reliability and trustworthiness are often questioned by sceptics*. This further justifies the need for the recommendations presented in this dissertation on a more fundamental level. To support this activity, an accessible and user-friendly online tool for microgravity analysis is presented that supports conventional data formats output by commercial CG-5 and CG-6 gravimeters [*Koymans*, 2022a]. The implemented methodology (appendix A) is suggested as the most effective approach to treat campaign gravity surveys that are routinely completed on volcanoes worldwide (**Research Question III**).

> **Research Question III** – What is the optimal strategy for the collection and treatment of campaign microgravity data to avoid the introduction of additional, avoidable uncertainties?
>
> Chapters 4 and 5 of this dissertation

Finally, although campaign and continuous microgravity observations provide additional constraints on subsurface mass processes, it should not be forgotten that volcano monitoring requires a multi-disciplinary approach. Acknowledging that, which is perhaps also the greatest shortcoming of this dissertation, it is evident that in order to fully understand the processes that initiate, sustain, and terminate volcanic eruptions data and knowledge from different sources and geoscientific communities are required. These observations may for example come from other geodetic measurements such as as ground tilt, deformation from GNSS and InSAR, precise leveling. But should also include volcano seismology (including tremor, long-period, and volcano-tectonic events), acoustics, gas emissions, and thermal and radar observations. In volcano monitoring infrastructures, such a diverse set of available geophysical data sources clearly expresses the desire for a shared data infrastructure. For this reason, an effort was made to integrate data from the absolute quantum gravimeter that was deployed in the NEWTON-g gravity imager with the European Integrated Data Archive (EIDA), an existing FAIR data infrastructure. This integration is discussed in the concluding chapter 6 and will facilitate the reuse of microgravity data. This chapter also serves as the synthesis of the dissertation with conclusions and key findings from the dissertation, identified shortcomings, as well as recommendations for future work.

*Including the author of this dissertation.

2

Performance assessment of geophysical instrumentation through the automated analysis of power spectral density estimates

Abstract This chapter describes an automated data quality verification procedure supported by a database of power spectral densities (PSD) estimates for geophysical waveform data. The Royal Netherlands Meteorological Institute (KNMI) manages a 100 TB archive of continuous geophysical data collected from accelerometers, geophones, broadband seismometers, and infrasonic arrays deployed across the continental and Caribbean Netherlands. This rapidly expanding network at a scale of over 700 instruments makes the manual evaluation of data quality impractical and must be supported by data policies and automated methods. A technique is presented to compress and store PSD estimates in a database with a storage footprint of less than 0.05 % of the raw data archive. Every week, the instrument performance is validated by comparing statistical properties of its latest monthly probabilistic PSD distribution to strict quality metrics. The criteria include thresholds based on global noise models, datalogger quantization noise models, constraints imposed by ambient noise conditions, and confidence intervals based on PSD estimates calculated from validated archived data. When a threshold is crossed, the station operator is alerted of the suspected degraded instrument performance, severely limiting the required amount of manual labor and associated human errors. The automated PSD assessment technique is applicable to accelerometers, geophones, broadband seismometers, infrasonic stations, and is demonstrated to be extendable to hydrophones, gravimeters, tiltmeters, and GNSS receivers. The approach is therefore suitable for other

Published as: Koymans et al., 2021. Performance assessment of geophysical instrumentation through the automated analysis of power spectral density estimates. Earth and Space Science, 8(9), doi: 10.1029/2021EA001675.

geophysical monitoring infrastructures, e.g., observational networks dedicated to continuous volcano monitoring. Making it possible to detect degraded instrument performance that may otherwise remain undetected.

2.1 Introduction

Geophysical monitoring infrastructures passively observe the effects of natural phenomena that continuously take place and evolve in the atmosphere, oceans, and the solid earth. At our disposal is a set of unique and complementary instruments and sensing techniques that each provide a piece of key insight towards understanding these processes. The different strengths and sensitivities of each different geophysical sensor and measurement contributes its part in a larger observational network. Because of this wide variety of sensors and data types, it is beneficial to find one universal data quality assessment technique that is effective for all instruments leveraged in a geophysical operational network.

The Royal Netherlands Meteorological Institute (KNMI) operates different types of geophysical instruments that are deployed in the Netherlands Seismic-Acoustic Network (NSAN), both in the continental Netherlands [*KNMI*, 1993] and Carribean Netherlands [*KNMI*, 2006]. The observational network (fig. 2.1) comprises over 700 instruments including geophones, accelerometers, broadband seismometers, infrasonic arrays, and four GNSS receivers. The monitoring capability of the network is promptly expanded when new observations are required, or new sensing techniques become more affordable and accessible.

Most geophysical monitoring infrastructures are designed for the assessment of hazards that include e.g., natural and induced earthquakes [*Camelbeeck and van Eck*, 1994; *Spetzler and Dost*, 2017], explosions [*Ruigrok et al.*, 2019], infrasonic events [*Evers et al.*, 2007], observations of volcanic unrest [*Carbone and Greco*, 2007; *Sparks et al.*, 2012], and to verify compliance with the Comprehensive Nuclear-Test-Ban Treaty [*Coyne et al.*, 2012], followed by rapid dissemination of information to the public. With advances in sensing techniques and rapidly expanding network densities due to rising interest in e.g., geothermal exploration and large-N networks, data quantities are growing at an increasing rate (e.g., [*Dost*, 1994; *Strollo et al.*, 2020]). As mandated by data volumes easily exceeding hundreds of terabytes across data centers and growing beyond tens of terabytes per year, quality verification of data must increasingly rely on automated, trusted, and documented policies and procedures.

Obtaining reliable data from geophysical monitoring infrastructures begins with adequate network design, proper instrument installation and configuration, completeness of metadata, and setting up the necessary infrastructure for reliable telemetry to transfer the recordings to a data center for storage. From there, continuous quality assessment of the incoming data is the backbone to sustain consistent and high data quality standards and create reliable products. The challenges in maintaining an extended and diverse network of geophysical instruments are to verify that 1) the configuration of each instrument in the field is consistent with its metadata, and 2) the performance of the instruments does not degrade over time. Developments towards (automated) geophysical data quality monitoring systems have been made over the last decades, e.g., PQLX [*McNamara and Boaz*, 2006a], MUSTANG [*Ahern et al.*, 2015], DQA [*Ringler et al.*, 2015], WFCatalog [*Trani et al.*, 2017], AutoStatsQ [*Petersen et al.*, 2019], and component ratio monitoring [*Pedersen et al.*, 2020]. These systems are often designed to detect and identify data anomalies due to changes in e.g., local site conditions, technical instrumental problems, timing issues, transmission failures, and to reveal inconsistencies in the instrument metadata. Because these problems are inherent to all sensor deployments and exist independently from the instrument in question, similar data quality assessment techniques can be applied for different instruments. Often such anomalies and discrepancies are detected by a researcher, for example after analyzing a series of earthquake data. A more proactive approach is favorable, and this study presents a system that facilitates the discovery of such potential problems at an early stage based on the comparison of instrumental power spectral density (PSD) estimates against strict quality metrics. Our work extends previous developments in the analysis of PSD estimates for the purpose of quality control by introducing multiple new independent metrics and making recommendations on what metrics to apply to various instruments in a geophysical monitoring network.

The performance of many geophysical instruments can be assessed through a PSD estimate of a segment of its output recording [*Rosat et al.*, 2004; *McNamara and Boaz*, 2006b]. This estimate is a measure of how the power density of the signal is distributed over the full instruments bandwidth. Using estimates of the PSD from different time segments, a probability density function (PPSD) can be aggregated, which introduces a confidence interval on the stochastic PSD estimates. The performance of an instrument can be monitored through the temporal evolution of the PSD, for example at discrete selected frequencies [*De Zeeuw-van Dalfsen et al.*, 2018], or over its full bandwidth. For example, the PSD estimates of broadband recordings of ground motion are expected to fall within a specific range e.g., the New High (NHNM) and New Low Noise Models (NLNM) derived from global observations [*Peterson*, 1993]. Besides the NHNM and NLNM, alternative statistical bounds are described in the literature e.g., [*Berger et al.*, 2004] and [*Castellaro and Mulargia*, 2012]. Similar confidence intervals have been estimated for accelerometers [*Cauzzi and Clinton*, 2013] and infrasonic sensors and hydrophones [*Brown et al.*, 2012]. These models can serve as a preliminary constraint to verify that the waveform data is recorded within prevalent environmental conditions. Because variations in the PSD can possibly be attributed to local underlying geophysical processes, [*García et al.*, 2006; *Burtin et al.*, 2008] global models are often insufficient, and every instrument and site needs to be verified independently. Furthermore, the global models themselves are sensitive to uncertainties introduced by the PSD processing and smoothing that may exceed the uncertainty of the data [*Anthony et al.*, 2020]. Therefore, it is recommended that a combination of multiple strategies is used to automatically monitor the performance of geophysical instrumentation when using PSD estimates.

For each instrument constraints are defined on the statistical parameters of the PPSD calculated from the latest month of data based on 1) global noise models from the literature, 2) datalogger quantization noise models following *Bennett* [1948] and *Sleeman* [2006], 3) confidence intervals established on manually validated archived waveform data [*McNamara et al.*, 2009], and 4) regionally expected ambient noise characteristics. Datalogger quantization noise models are derived for the network of geophone boreholes and accelerometers in the Dutch province of Groningen where the instruments are configured to measure relatively large ground accelerations without distortion or clipping. As a consequence, the self-noise of the equipment, which is dominated by digitizer noise, comes close to and at certain frequencies dominates over the ambient noise level. For each sensor and datalogger combination, a model for digitizer quantization noise is derived that is used as a theoretical lower limit of the PPSD that can be observed. In addition, for instruments that are expected to be dominated by digitizer noise at low frequencies, the PPSD distribution is expected to fall close to the derived quantization noise level. It is demonstrated that the approach is easily applicable to tiltmeters, despite not being incorporated in the NSAN operational network.

Confidence intervals on the PPSD of incoming data can be derived from an upper and lower percentile of the PPSD of archived and validated data from the same instrument recorded in regular conditions over a full year to encapsulate seasonal variability. These confidence bands represent local low and high noise models that are distinct for each instrument and site. Additionally, when an instrument records more than just instrumental noise, i.e., environmental noise, a scatter is introduced on the stochastic PPSD distribution that is confirmed to be present in the data. Finally, the characteristics of the microseisms are used. In the Netherlands, the spectral peak of the microseisms provides a stable and well-recognizable regional reference for performance monitoring and is used to constrain characteristics of the PSD estimates.

This study focuses on the NSAN, but also demonstrates the approach for geophysical instrumentation that are not presently operated in the monitoring network. In the verification process that is operated weekly, statistical parameters of the latest monthly PPSD of all instruments are compared to these constraints. When an instrument fails to meet these criteria, the station operator is notified about the potential degraded performance of the equipment. The methodology and background to calculate the PSD estimates are described, followed by the database design which is optimized for the storage and retrieval of many PSD estimates. Next, the spectral constraints devised for accelerometers, geophones, broadband seismometers, microbarometers, gravimeters, tiltmeters, hydrophones, and GNSS receivers are discussed, that are suggested to be applied as quality metrics. Finally, the quality verification procedure operated by the NSAN and its process logic are described, with examples of degraded instrument performance detected by the system. This study is concluded with an operational product for automated data quality assessment for geophysical instruments deployed in the NSAN.

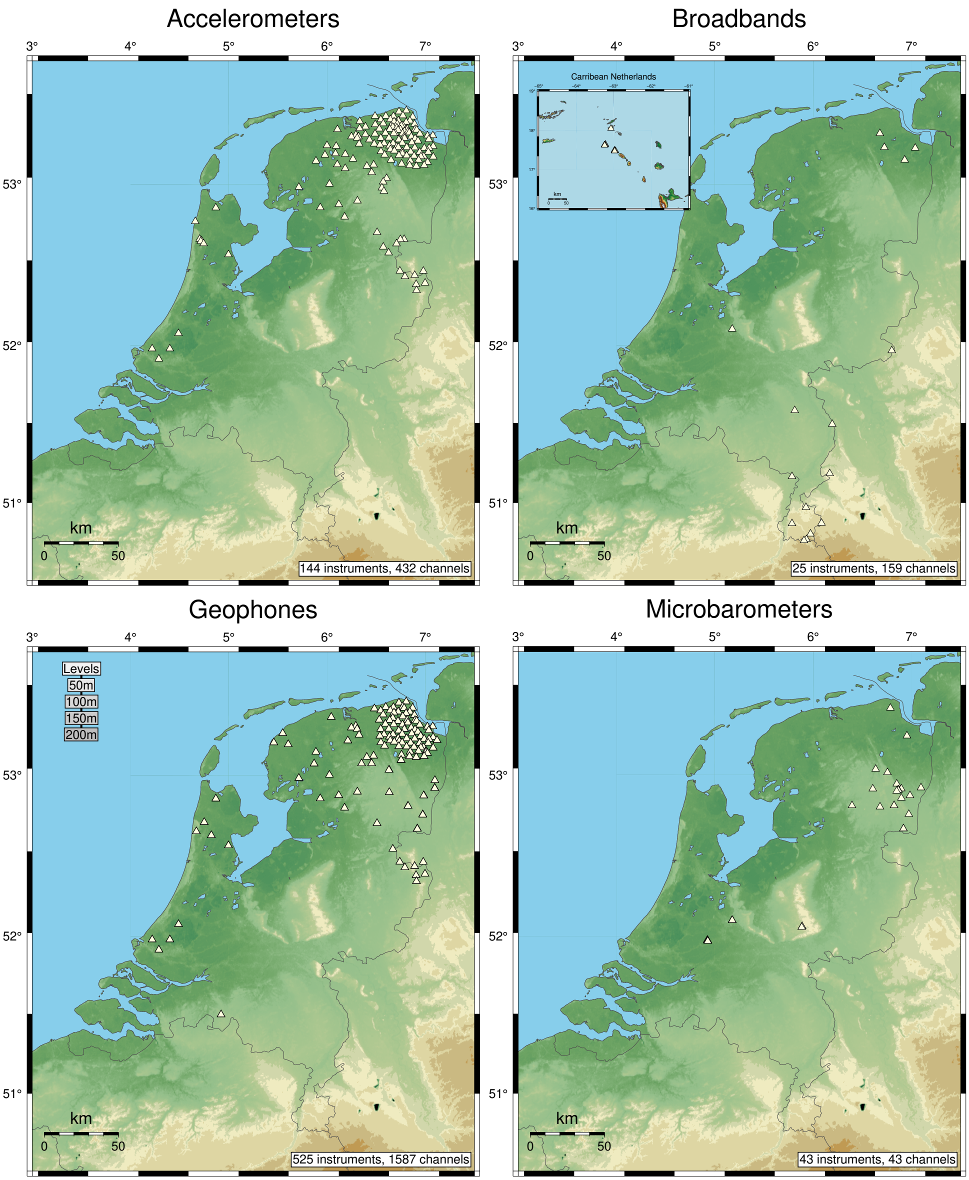

Figure 2.1: *Map of the (Caribbean) Netherlands showing accelerometers, broadband seismometers, geophones, and microbarometers deployed in the NSAN. The network is dense near the Groningen gas field in the north of the Netherlands where geophones are deployed in* 200 m *deep boreholes at* 50 m *depth intervals. In October 2020, the network consists of a total number of 728 instruments. Not illustrated are four GNSS receivers installed on the islands of St. Eustatius and Saba in the Caribbean Netherlands.*

2.2 Methodology

2.2.1 Definition of the PSD Estimate

The PSD is calculated by taking the Discrete Fourier Transform (DFT) of a time-variant signal and squaring the magnitude of its complex coefficients, normalizing the power per unit frequency following e.g., *Heinzel et al.* [2002]:

$$PSD(f_\mathrm{m}) = 2 \cdot \frac{|y_\mathrm{m}|^2}{\mathrm{N} f_\mathrm{s}} \tag{2.2.1}$$

where f_s represents the instrument sampling frequency, $|y_\mathrm{m}|$ the complex modulus of the m^{th} frequency component of the DFT, and N the number of samples used in the DFT. In this convention, the power distribution is mapped to positive frequencies, expressed by the factor 2.

2.2.2 Systematic Computation of Power Spectral Densities

Despite the publication of the Fast Fourier Transform (FFT) [*Cooley and Tukey*, 1965], the calculation of power spectra remains a computationally intensive task. It is inefficient to load archived data from disk and calculate PSDs on demand for instrumental monitoring purposes and a caching strategy for the PSD segments is recommended. The computational challenge is then shifted towards the storage of millions of spectra and their efficient retrieval. Our solution is to store a smoothed and decimated approximation of the PSD estimates using a custom compression scheme that minimizes the storage footprint and offers a flexible and fast retrieval mechanism.

PSDs are estimated from hourly segments of continuous data with 50 % overlap between segments. All data are preprocessed using the Python library ObsPy [*Beyreuther et al.*, 2010] by which overlapping samples in one segment are removed and time-discontinuous data segments are rejected from further processing. The approach of *McNamara and Buland* [2004] is followed and PSDs are calculated within a predefined fixed frequency range. During the calculation, the hourly waveform segment is split into thirteen segments of even length with 75% overlap. For each partial segment, the linear trend is removed and a normalized 10 % cosine taper is applied to dampen the effect of spectral leakage caused by discontinuity at the segment edges. The instrument response is evaluated using ObsPy through *evalresp*, and corrected for in the frequency domain to obtain spectral amplitudes in physical units (e.g., $\mathrm{m\,s^{-2}}$ for seismic data). The spectral amplitudes are squared to power and normalized following eq. (2.2.1). The thirteen resulting PSDs are averaged to obtain the final PSD for that hourly segment following the method of *Welch* [1967]. The averaged PSD of the hourly segment is smoothed over a full octave band and is expressed in dB and stored for 256 fixed frequencies at one-eight octave intervals, following the sequence defined in eq. (2.2.2), starting at $f_0 = 1024\,\mathrm{Hz}$ for all instruments. A property of this sequence is that it generates whole integer frequency bin centers at e.g., 1, 2, 4 Hz commonly used for inspection.

$$2^{-0.125n} \cdot f_0 = [f_0, \ldots, 4, \ldots, 2, \ldots, 1 \ldots, f_n] \tag{2.2.2}$$

The upper and lower bounds of a frequency bin can be calculated by adding or subtracting half the bin width from its center frequency respectively. Powers for frequencies above the Nyquist frequency and below the limit that is imposed by the segment length used in the DFT are registered as invalid by a reserved value. The number of selected fixed frequencies in eq. (2.2.2) can be chosen arbitrary, but must be large enough to include the low frequencies of interest. The remaining power values are rounded to 1 dB sized bins. The smoothed and decimated approximation of the PSD segment reduces its compressed storage size significantly, and more optimization can be achieved with an appropriate data storage strategy as defined below.

2.2.3 Storage Strategy

For each segment, the result relates frequency to power at discrete bins that can be represented as a sparse two-dimensional binary matrix, which shape depends on the chosen binning granularity (fig. 2.2). Because the PSD(f) relationship is continuously spaced across the frequency domain, and the chosen frequency and amplitude bins are equivalent for all segments, the full spectrum can thus be represented as a contiguous one-dimensional array of PSD values, with the associated frequencies implicitly following the sequence defined in eq. (2.2.2).

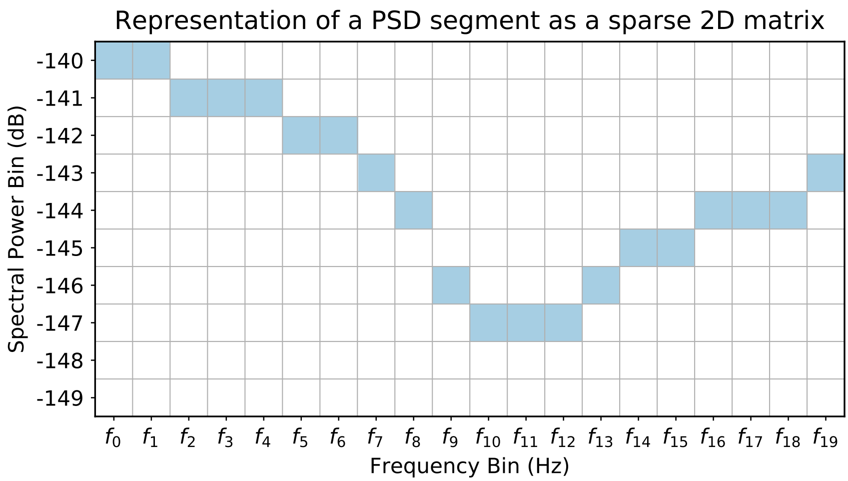

Figure 2.2: *Synthetic PSD estimate over 20 arbitrary frequency bins. The binned frequencies and corresponding power densities can be treated as a sparse 2-dimensional binary matrix. This particular segment only requires a total of 20 bytes to be stored, outside of additional metadata required for full reconstruction.*

Invalid powers for frequencies that are beyond the Nyquist frequency are eliminated from the start of the array, and the index of the first valid frequency (f_{offset}) is stored auxiliary to the array so that the spectrum can be reassembled. Power values for low frequencies beyond the segment length will truncate the array to the lowest valid frequency and simply reduce its length. Most currently available recordings are digitized with 24 bits, which results in a dynamic range that is smaller than 150 dB. When the power densities are rounded to whole integer dB units the difference between the minimum and maximum value fits within a single byte integer. This property enables a shift of the PSD by an integer PSD_{shift} that brings all power values to the positive range, and is stored as metadata to the array. A single PSD segment can thus be represented as an array of 8-bit unsigned integers composed as follows:

$$\text{UInt8Array} = [PSD_{f_{\text{offset}}},\ PSD_{f_{\text{offset}+1}},\ \ldots,\ PSD_{f_{\text{offset}+N}}]_{\text{uint8}} \tag{2.2.3}$$

Where N represents the number of valid frequencies that is a function of the instrument Nyquist frequency and thus the sampling interval. Following this method, the PSD of a single segment, even at high sampling rate, can be represented with a small storage footprint (fig. 2.3). This size excludes the size of the metadata required to identify the segment within the database, including the offset (f_{offset}) and shift (PSD_{shift}) required for reconstruction.

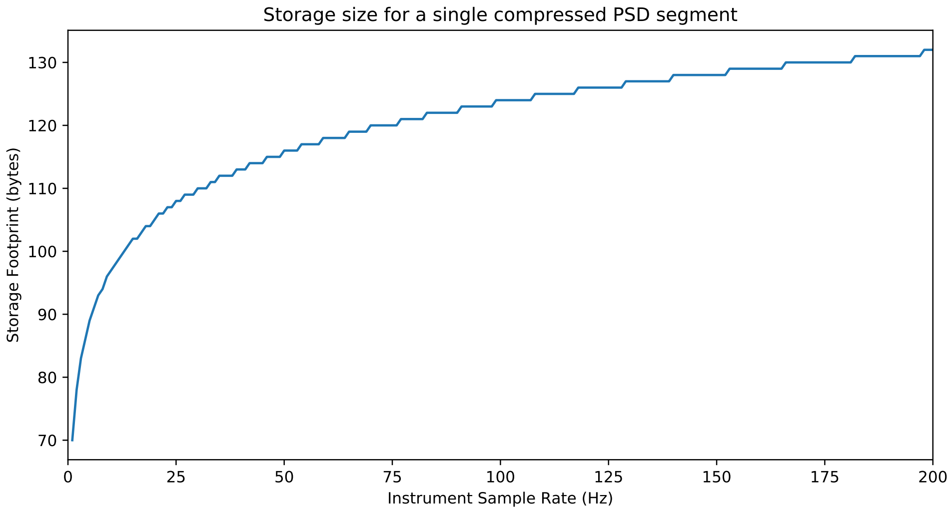

Figure 2.3: *Showing the number of bytes required to store a single PSD array binned at* 1 dB *and one-eight octave bins within the fixed frequency range defined in eq.* (2.2.2)*. The number of valid frequencies is limited by the Nyquist frequency of the instrument and the chosen segment length for the DFT. The relationship between the instrument sample rate and required storage is logarithmic.*

2.2.4 Database Selection

The derived PSD segments are stored in a database with metadata that describes its time coverage and the seismological standard SEED data stream identifier [*Ahern et al.*, 2007]. This metadata is indexed and used to retrieve the spectra that satisfy particular search criteria (e.g., a specific instrument and time span). A checksum of the raw waveform data and metadata are included to detect changes in the (meta)data that mandates the reprocessing of PSD segments in the database.

Several technical solutions were explored for the storage of millions of PSD segments, accounting for the way the database will be used. Due to the inherent variability of individual PSD estimates, the PPSD is used to represent a statistical distribution of multiple PSD estimates. The objective of aggregating individual binned PSD segments to a PPSD is equivalent to generating a two-dimensional histogram (PSD, f) that is accumulated over a third temporal dimension (t), namely the PSD segments over time. Initially, SciDB was tried [*Stonebraker et al.*, 2013] as it is designed for array and matrix operations of this kind. One large challenge was encountered as arrays are expected to have a predefined size, which conflicted with a continuously growing number of PSD segments over time. The ObsPy library offers a method to calculate PPSDs and save them using NumPy [*Van Der Walt et al.*, 2011] compressed files. This approach has a storage footprint that is orders of magnitude larger than our custom compression scheme and was therefore not used. The NoSQL database MongoDB became the storage database of choice, as the schema-less nature of this database facilitates rapid prototyping and it quickly evolved into a functional product. Alternatively, a RDBMS (e.g. MySQL, MariaDB, PostgreSQL) can be chosen with no significant effect on the database performance. The compression scheme presented in this paper keeps the spectral database of the current waveform archive (100 TB) at a manageable size of roughly 30 GB, which is about 0.03% of the raw data volume. As of 2022, the database has been upgraded to DynamoDB.

2.2.5 Aggregation to Probabilistic Power Spectral Densities

On top of the spectral database an Application Programming Interface (API) was developed to find and access the PSD estimates and aggregate the results. When an API request is made, the PSD segments that match the request criteria (e.g., temporal coverage, or that for a particular SEED identifier) are returned. These segments are passed to a subroutine that allocates an empty zero-filled matrix of dimensions 255 × 256 ($\mathrm{N_P} \times \mathrm{N_f}$) that corresponds to the number of selected power and frequency bins respectively. For every PSD segment, the array of 8-bit unsigned integer values are unpacked and used to find the respective cells within the matrix to be incremented, where invalid values are discarded. The frequency bin can be calculated from its frequency offset (f_{offset}), in combination with its particular index in the array. The power value is represented by the unsigned byte value corrected for by the power shift (PSD_{shift}) value. When all spectra have been aggregated the resulting matrix is divided element-wise by the total number of segments used to convert the absolute values to probability of occurrence. The resulting histogram then represents a fully reconstructed PPSD for the requested options and can be returned to the client for further analysis.

2.3 Automating PSD Quality Control for the NSAN

Extensive and precise quality control of data is one of the most vital and resource intensive tasks of a data center. The large quantity of incoming geophysical data from the NSAN requires that the process of quality control becomes increasingly automated. The operational procedure includes all types of instruments and divides them into three distinctive processing stages. Stage 0) includes instruments in the first month of operation and serves to assert whether the instrument performs well enough to be moved into production. Stage 1) contains the set of instruments in the first year of operation until one year of high quality data is available and the instrument proceeds to the following stage. Stage 2) enforces the guarantee that the quality of an instrument does not degrade over time by comparing incoming data to a history of manually validated data that was archived during stage 1 of the instruments deployment.

2.3.1 Instrument Performance Criteria

Power spectra can be utilised as a quality metric as the distribution of power over frequency generally falls within an expected range depending on ambient noise conditions and intrinsic instrumental noise. In the following sections the recommended PPSD criteria for many geophysical instruments are defined and discussed that can be applied as a performance metric. In the NSAN, the latest monthly PPSD of all instrumental channels are automatically verified in weekly intervals against these criteria. If one of the metrics fails, the station operator is notified of the suspected degraded instrument performance. Geophysical instruments that are not part of the NSAN operational network are discussed and it is demonstrated how these instruments can be easily integrated in the future.

Quantization Noise Constraint

Geophysical monitoring infrastructures consist of many different types of sensors and dataloggers, and the combination of the two is referred to as an instrument. The dynamic range of an instrument is determined by the ratio of the maximum to minimum amplitude of a signal that can be recorded and can be limited by either the sensor or datalogger [*Steim*, 2015]. The maximum amplitude is usually defined by the clip level of the sensor, whereas the minimum amplitude is often limited by self-noise of the digitizer. An expression for the self-noise of the system thus provides a constraint on the minimum PSD level that can be expected in any output recording.

For ground motion data, the instrumental noise is expressed in terms of ground acceleration power density ($\mathrm{m^2\,s^{-4}\,Hz^{-1}}$). The maximum sensor amplitude (i.e., clip level) A must thus be expressed in terms of ground acceleration too. This amplitude is constant for accelerometers, i.e., a flat frequency response, and frequency dependent for e.g., geophones and broadband seismometers (fig. 2.4). The frequency band for the quantization noise constraint is defined between 0.033 Hz and 80 % of the instrument Nyquist frequency. The self-noise models are inaccurate near the Nyquist frequency where the effects of anti-aliasing filters play a significant role in decreasing the instrument sensitivity [*Sleeman*, 2006], and at low frequencies due to instabilities caused by tilt, wind, pressure, and temperature variations.

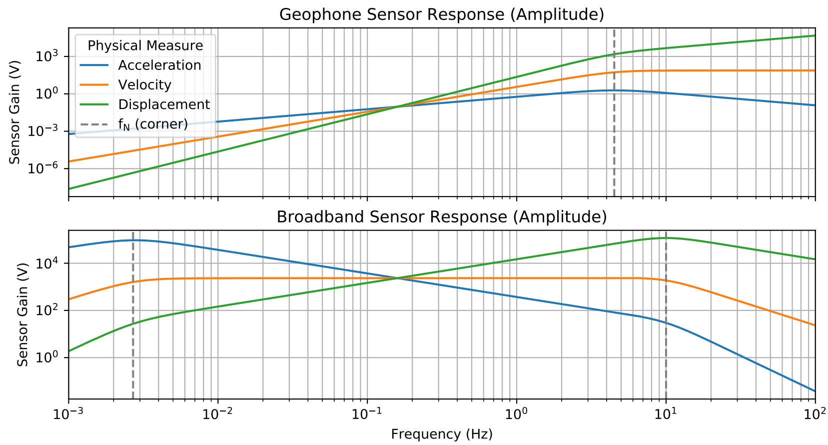

Figure 2.4: *Examples of frequency-dependent sensor gain. Top) sensor gain in* V *for a geophone with a resonant frequency of* 4.5 Hz*, a damping factor of 0.702, and an effective generator constant* $\mathrm{G_e}$ *of* $75.8\,\mathrm{V\,m^{-1}\,s^{-1}}$*. The amplitude of the frequency response to ground displacement (green), velocity (orange), and acceleration (blue) is shown. The sensitivity of geophones to ground acceleration decays around its resonant frequency* f_0 *indicated by the vertical grey line. Bottom) sensor gain for an STS-1 broadband station with a* $\mathrm{G_e}$ *of* $2332\,\mathrm{V\,m^{-1}\,s^{-1}}$*, corner frequencies of* $f_0 = 360\,\mathrm{s}$ *and* $f_1 = 10\,\mathrm{Hz}$*, and damping factors* $h_0 = 0.707$ *and* $h_1 = 0.623$*. The broadband station is much more sensitive to ground displacements at low frequencies compared to a geophone. An accelerometer has a flat response to ground acceleration below its resonant frequency and is therefore not shown.*

Analog-to-digital converters (ADC) discretize continuous functions to quantization levels, and in this process truncation errors to the nearest quantization level are introduced. *Bennett* [1948] derived an equation for the quantization noise power of an ADC over a full-load sine wave:

$$\epsilon_{\mathrm{rms}}^2 = \frac{\Delta^2}{12} \tag{2.3.1}$$

where Δ is the digitizer resolution or quantization interval which can be expressed as $2A/2^n$ with n the number of effective bits of the digitizer, and A the full-load amplitude. Despite the assumption of a full-load sine wave input, this equation appears to hold for sufficiently complex signals passing over many quantization intervals [*Oppenheim and Schafer*, 2009]. For a white noise spectrum the power density is constant over the full bandwidth. This means that the PSD of the quantization error can be estimated as the mean square error $\epsilon^2_{\mathrm{rms}}$ derived in eq. (2.3.1) replaces $|y_{\mathrm{m}}|^2$ over N in eq. (2.2.1) following Parseval's theorem:

$$PSD_{\mathrm{min}} = 2 \cdot \frac{\epsilon^2_{\mathrm{rms}}}{f_{\mathrm{s}}} \tag{2.3.2}$$

eq. (2.3.2) can be expanded using eq. (2.3.1), rewritten, and expressed in dB relative to a reference value of $1\,\mathrm{m}^2\,\mathrm{s}^{-4}\,\mathrm{Hz}^{-1}$ for ground motion instruments.

$$PSD_{\mathrm{min}} = 10\log_{10}\left(\frac{T}{6}\left(\frac{2A}{2^n}\right)^2\right) \tag{2.3.3}$$

eq. (2.3.3) describes an absolute minimum white noise spectrum that is uniform over all frequencies. This level depends on the sampling interval T (f_s^{-1}), as the total noise power does not change with a different number of samples per second, effectively reducing the power spectral density [*Sleeman*, 2006; *Oppenheim and Schafer*, 2009].

Inside active electronic components the electronic noise level is inversely proportional to frequency. This f^{-1} type of noise dominates the power spectrum at low frequencies. Following *Sleeman* [2006], $PSD_{\mathrm{min}}(f)$ is expressed, now a function of frequency, as a superposition of the flat white noise spectrum defined in eq. (2.3.3) and a frequency dependent pink noise spectrum. This model assumes that the frequency dependent noise is thus proportional to f^{-1}, which often holds for instruments in the NSAN, but is not a universally valid assumption under all circumstances.

$$PSD_{\mathrm{min}}(f) = 10\log_{10}\left(\frac{T}{6}\left(\frac{2A}{2^{n_1}}\right)^2 + \frac{T}{6f}\left(\frac{2A}{2^{n_2}}\right)^2\right) \tag{2.3.4}$$

where n_1 and n_2 represent the effective number of bits for each spectrum respectively. From experience with instruments deployed in the NSAN it is found that the value of n_2 is related to n_1 following $n_2 = n_1 + 1$. This eliminates one unknown and combines both number of effective bits into a single variable number of proxybits η. When the sampling interval T is also implicitly included in the number of unknown proxybits η, eq. (2.3.4) can be reduced and rewritten to:

$$PSD_{\mathrm{min}}(f) = 10\log_{10}\left(\frac{1}{6}\left(\frac{2A}{2^{\eta}}\right)^2\left(1 + \frac{1}{4f}\right)\right) \tag{2.3.5}$$

For every datalogger and sensor combination the single unknown variable η is estimated by fitting it to an instrument that is dominated by self-noise over part of its bandwidth, generally at frequencies below 1 Hz where f^{-1} noise dominates. The exact value of η must be close to the number of bits provided by the digitizer manufacturer but depends on the (noise) specifications of the instrument and the configured internal (over)sampling interval of the datalogger. This constraint is determined once, and then applied to other instruments of the same type. The minimum of the PPSD may under no circumstances fall below this theoretical baseline. The results for all types of instruments in the NSAN are compiled in table 2.1, and details on the methodology per instrument type is discussed in section 2.3.2.

Instrument	A (m s^{-2})	V_{sensor}	$V_{\text{datalogger}}$	η
Batch-1 Accelerometers	4g	±20	±20	24.7
Batch-2 Accelerometers	2g	±5	±20	22.7
Batch-3 Accelerometers	2g	±20	±20	24.7
Batch-4 Accelerometers	2g	±5	±5	24.7
Etna-2 Accelerometers	2g	±2.5	±2.5	24.5
SM6 Geophones	f-dependent	±2.5	±2.5	24.3
SM6H Geophones	f-dependent	±2.5	±2.5	24.3

Table 2.1: *Compilation of different accelerometers and geophones deployed in the NSAN and the parameters used to calculate the instrumental lower noise bounds. All instruments sample at* 200 Hz*. A represents the maximum amplitude that may be frequency dependent, V_{sensor} and $V_{\text{datalogger}}$ the sensor output voltage and datalogger input voltages, respectively. The difference in the estimated proxybits between batch-2 and batch-3 accelerometers emerges from a different setting between the configured sensor output voltage range between* -5 V *to* 5 V *and expected digitizer input voltage between* -20 V *to* 20 V*. This inconsistency in the instrument configuration introduces a range of factor four that is never digitized, effectively not using two available bits on the datalogger.*

Global Noise Model Constraint

For highly sensitive instruments (e.g., broadband seismometers and gravimeters), the level of ambient noise usually exceeds that of quantization noise, and a generic global noise model is more suitably chosen as a lower limit. One example is the *Peterson* [1993] global noise model that includes long period disturbances, microseisms, and anthropogenic noise. Instruments installed in a high-noise environment may produce PSDs above the NHNM, whereas the PSD of instruments installed in quiet ambient conditions sites may fall near the NLNM. These models are based on observations from broadband instruments and therefore the NLNM does not translate well to strong motion accelerometers [*Cauzzi and Clinton*, 2013] and geophones, of which many are installed in the NSAN. The International Data Centre (IDC) provides similar noise models for acoustic and hydroacoustic data [*Brown et al.*, 2012]. The median of the monthly latest PPSD of an instrument should fall within its respective global noise model from the literature and constrains the PPSD over a large bandwidth.

Microseisms Constraint

Oceanic wave-wave interaction and coastal swell impose a distinctive and reliable noise field over the Netherlands that is limited to a certain frequency band [*Kimman et al.*, 2012]. These microseisms serve a stable minimum and maximum constraint on the expected level of ambient noise at the instrument site. The geographical extent of this field is frequency dependent, but frequencies of $f \approx 0.3\,\mathrm{Hz}$ attenuate less than higher frequencies and thus can be observed consistently across the Netherlands, even at depths down to 200 m depth by geophones in boreholes (fig. 2.5). The minimum and maximum expected power at $f \approx 0.3\,\mathrm{Hz}$ was constrained over a full year at monthly intervals with 3σ confidence intervals for various distances from the coast as illustrated in fig. 2.6.

Despite a clear relationship between the observed power spectral density around the secondary microseism frequency and distance to an active coastal area can be recognized, a single minimum and maximum for the entire Netherlands is used. The expected noise power in the Netherlands at $f \approx 0.3\,\mathrm{Hz}$ appears limited between $PSD_{\max}$ at $-90\,\mathrm{dB}$ and $PSD_{\min}$ at $-140\,\mathrm{dB}$. These two values provide a constraint for the expected level of the microseism at 0.3 Hz specifically for the Netherlands.

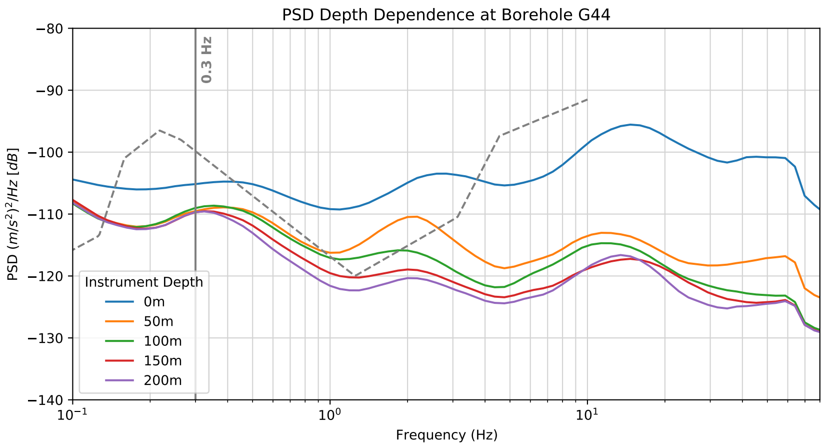

Figure 2.5: *Showing the averaged and smoothed PSD over 2019 of instruments G440 (surface accelerometer) and G441-G444 (geophones) deployed at* 50 m *depth intervals respectively. Noise at higher frequencies attenuates faster than lower frequencies with increasing depth. The noise power observed at all depth levels is roughly equivalent at* 0.3 Hz*. The surface accelerometer is influenced by tilt and atmospheric variations and always expresses higher noise levels when averaged over a full year. The NHNM after [Peterson, 1993] is shown for reference.*

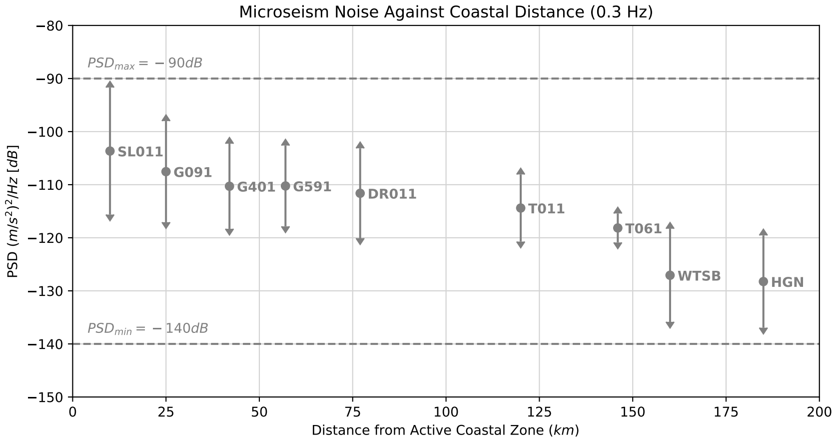

Figure 2.6: *The PSD at 0.3 Hz, as a function of distance from the (active) coast. The round markers indicate the average levels over one year of data from the labeled station, the bars illustrate the* -3σ *and* $+3\sigma$ *region (99.7% confidence region). From this distribution of PSD levels, the microseism constraint is derived, yielding at* $0.3\,\mathrm{Hz}$ *a* $PSD_{\min}$ *of* $-140\,\mathrm{dB}$ *and a* $PSD_{\max}$ *of* $-90\,\mathrm{dB}$.

Low Frequency Constraint

This threshold is suggested for accelerometers and geophones at $f = 0.025\,\mathrm{Hz}$ where the PSD is expected to be dominated by quantization noise. As the probabilistic distribution over these frequencies is narrow in this range, it is expected that the median of the power spectrum falls within the theoretically derived $PSD_{\min}$ from eq. (2.3.5) and $PSD_{\min} + 10\,\mathrm{dB}$.

Minimum - Maximum Difference Constraint

This threshold is introduced in order to detect broken sensors, or instruments for which the sensor is not (properly) connected to the digitizer. In these cases only instrumental noise is recorded. In normal conditions, the difference between the statistical minimum (2.5^{th}) and maximum (97.5^{th}) percentile of the PPSD over a month at $f \approx 3\,\mathrm{Hz}$ must display at least a typical scatter of $5\,\mathrm{dB}$. This frequency is chosen because it falls in the bandwidth of anthropogenic noise where a large variance in the PSD estimates is expected. If this difference is smaller than $5\,\mathrm{dB}$, a warning is issued that the instrument might be dysfunctional. False positives of this constraint have been identified in the system for functional instruments operating in very quiet conditions.

Default Bound Constraint

The default bound comparison comprises a check against a constant upper and lower default bound of $-80\,\text{dB}$ and $-140\,\text{dB}$ respectively for seismic instruments. These wide bounds are based on best estimates for the Netherlands from experience and have no physical background but should identify large anomalies at any frequency.

Percentile Constraint

The percentile constraint has been previously applied by e.g., *Ringler et al.* [2015] and is the most effective metric that is placed on instruments that have been deployed for over a year and have an archived record of manually verified high-quality data. For each instrument, from archived PSD estimates stored in the database an upper (2.5^{th}) and lower (97.5^{th}) percentile curve is computed. These data-driven bounds represent confidence limits for incoming data, within average environmental and instrumental noise for that particular instrument and location. Every week, the median of the latest monthly PPSD is compared against these statistical constraints, and when a threshold is crossed, the deviation may be caused by degraded instrument performance over time. Crossing of the bounds may also be caused by the addition or removal of a strong persistent source of noise. The sensitivity of the detection can be tuned by choosing a particular percentile limit (fig. 2.7). Furthermore, both the mean or median of the PPSD can be used as a trigger, where using the median will decrease the sensitivity towards single large anomalies.

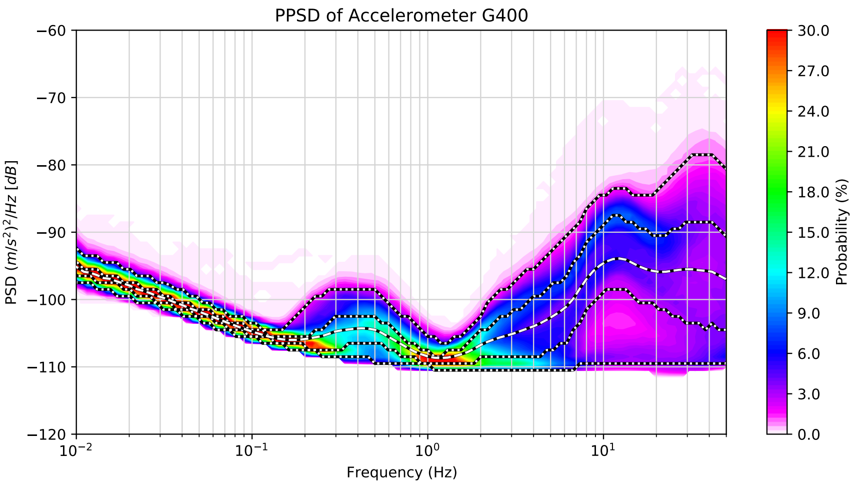

Figure 2.7: *Showing the PPSD of surface accelerometer G400 (vertical) with the* 50^{th} *(median; dashed),* 2.5^{th}, 25^{th}, 75^{th}, *and* 97.5^{th} *percentiles (dotted) projected on the probability density function.*

2.3.2 Quality Assessment of Geophysical Instruments

Accelerometers

Most accelerometers in the NSAN are of type EpiSensor (Kinemetrics) with a dynamic range of 155 dB and a clip level of 2 g or 4 g. The instruments sample ground acceleration at sampling rates of 200 Hz using 24-bit dataloggers. Accelerometers are characterised by a flat response to acceleration below their resonant frequency. Because the instrument is dominated by instrumental noise over part of its bandwidth, a lower limit imposed by quantization noise can be empirically derived. The maximum amplitude A expressed with respect to ground acceleration is thus equivalent to an accelerometers clip level, and can therefore be inserted in eq. (2.3.5) as a constant. The f^{-1} component is found by fitting eq. (2.3.5) to data to estimate the number of proxybits η as illustrated in fig. 2.8, resulting in an estimate of $\eta = 22.7$ for that particular instrument.

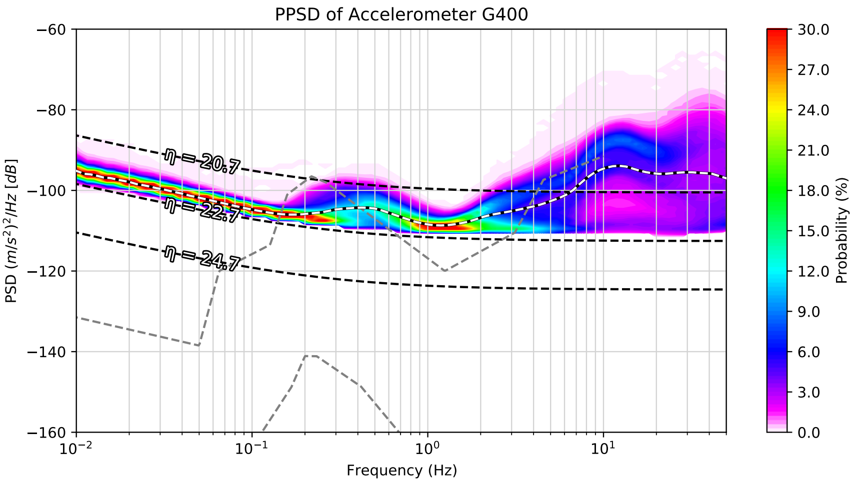

Figure 2.8: *PPSD of surface batch-2 accelerometer G400 (vertical). The median value of the PPSD is illustrated by the dotted white line. Curves for various proxybits η following eq. (2.3.5) are shown. For this example, a value for $\eta = 22.7$ is found and used to model the digitizer quantization noise of this instrument and all other instruments with the same setup (table 2.1).*

Geophones

Geophones are passive velocity transducers and generally have a flat velocity response above their resonant frequency f_0 (fig. 2.4). Below this frequency the sensitivity decays proportional to a nominal damping factor h as described by e.g., *Havskov and Alguacil* [2016] and can be expressed in terms of ground acceleration:

$$T_{\mathrm{a}}^{\mathrm{v}}(\omega) = \mathrm{G_e}\frac{-i\omega}{\omega_0^2 - \omega^2 + 2i\omega\omega_0 h} \tag{2.3.6}$$

where $\mathrm{G_e}$ is the effective instrument generator constant in $\mathrm{V/m\,s^{-1}}$, ω and ω_0 are the angular and resonant frequency respectively. The maximum amplitude A expressed as acceleration is found by dividing the maximum sensor output voltage V_{max} by the complex modulus of the frequency dependent sensitivity as defined in eq. (2.3.6):

$$A(f) = \frac{V_{\mathrm{max}}}{|T_{\mathrm{a}}^{\mathrm{v}}(\omega)|} \tag{2.3.7}$$

The frequency dependent amplitude $A(f)$ for geophones expressed in acceleration can be applied in eq. (2.3.5), and used to estimate η following an identical approach as for accelerometers. Examples for station T064 for various proxybits η are illustrated in fig. 2.9, where the number of $\eta = 24.3$ is empirically recovered.

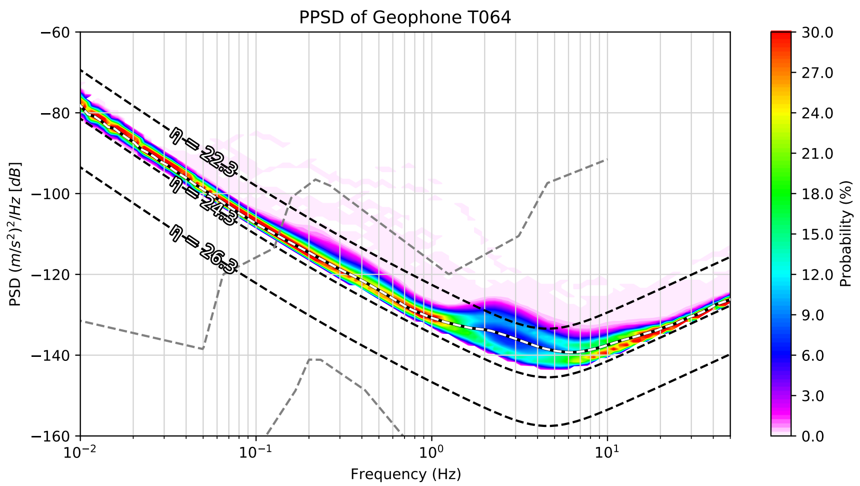

Figure 2.9: *PPSD of borehole SM6H geophone T064 (vertical) at* 200 m *depth. The median value of the PPSD is illustrated by the dashed white line. Curves for various proxybits η following eqs. (8), (9), and (10) are shown. For this example, a value of $\eta = 24.3$ is found and used to model the instrumental lower noise limit of this instrument and all others with the same setup (table 2.1).*

Broadband Seismometers

Broadband seismometers remain sensitive to ground accelerations at low frequencies, yet not to direct components at zero frequency (fig. 2.4). The STS-1 sensors installed in the network have a flat response to ground velocity between 360 s and 10 Hz. The high sensitivity of the sensor limits the maximum ground acceleration that can be measured by the seismometer before clipping. This suggests that for dataloggers with input that is a) aligned with the maximum sensor output without distortion, and b) have a large dynamic range ($> 140\,\mathrm{dB}$), may have quantization noise below the ambient noise (fig. 2.10). For an STS-1, because of its lower and upper corner frequency the response is expressed as eq. (2.3.6) multiplied by an extra term [*Dost and Haak*, 2002]:

$$T_{\mathrm{a}}^{\mathrm{v}}(\omega) = \mathrm{G_e}\frac{-i\omega}{\omega_0^2 - \omega^2 + 2i\omega\omega_0 h_0} \cdot \frac{\omega_1^2}{\omega_1^2 - \omega^2 + 2i\omega\omega_1 h_1} \tag{2.3.8}$$

where the indices represent the natural damping h_i and frequency ω_i of the lower (0) and upper (1) corner, respectively (fig. 2.4). The theoretical curves for two proxybits η are shown in fig. 2.10 and confirm that the limit falls below that of environmental noise. The presented equations are different for STS-2 [*Dost and Haak*, 2002] or STS-5 sensors and depend on the generation of the electronics, but the conclusion remains unchanged. A quantization noise model is not effective when it falls below a seismic noise model for the entire bandwidth of interest. For broadband seismometers the *Peterson* [1993] NLNM sets a stricter and more useful lower limit.

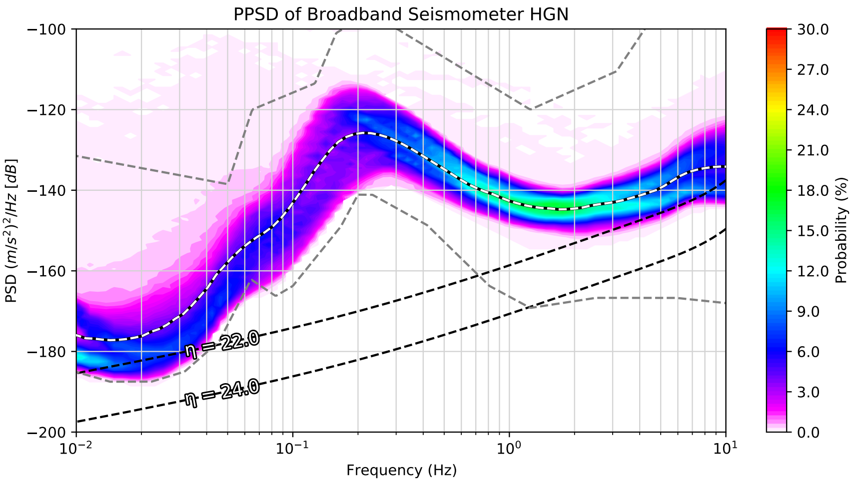

Figure 2.10: *PPSD of STS-1 broadband station HGN (vertical). The theoretical quantization noise of a datalogger for $\eta = 24.0$ is shown. From the data alone is it impossible to fit a model of quantization noise. Even at very low frequencies, electronic f^{-1} noise does not overtake the ambient noise signal. The quantization noise model has no added value, and a stricter measure for the expected lower limit of noise over the full bandwidth is the Peterson [1993] NLNM, illustrated in dashed grey.*

Microbarometers

The sensitivity of KNMI microbarometers [*Mentink and Evers*, 2011] to pressure is roughly flat throughout the entire frequency spectrum. The dynamic range of the instruments is limited by the sensor at 100 dB, while the 24-bit dataloggers yield a dynamic range in the order of 146 dB, prohibiting a fit for quantization noise. Furthermore, the ambient atmospheric noise also clearly exceeds the expected f^{-1} noise of 10 dB per decade, and thus a more reliable measure for the expected lower limit of noise over the full bandwidth is the IDC Infrasound Global High and Low Noise Models [*Brown et al.*, 2012] illustrated in fig. 2.11.

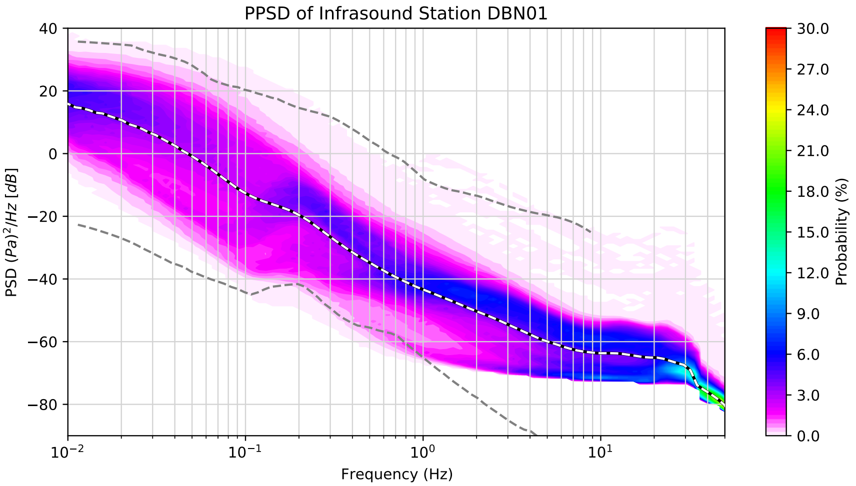

Figure 2.11: *PPSD of infrasound station DBN01 deployed at the Royal Netherlands Meteorological Institute for 2019 using a segment length of* 1 h *using a reference of* $1\,\mathrm{Pa}^2\,\mathrm{Hz}^{-1}$*. There is a large seasonal variability in the noise characteristics of the atmosphere, leading to a scattered PSD distribution. The IDC Infrasound Global High and Low Noise Models [Brown et al., 2012] are illustrated in dashed grey.*

Hydrophones

Hydrophones are not employed in the NSAN, and the PSD processing is done similarly to that of infrasonic stations, but the response of a hydrophone to pressure usually decreases with lower frequencies. It is challenging to determine an empirical noise model because the mechanical transfer functions of the instruments are more complex compared to seismometers. Instead, for demonstrative purposes the poles and zeros are evaluated from the metadata to obtain the precise instrument transfer function. However, this makes the analysis practically useless since one goal of this criteria is to detect inconsistencies with the instrument metadata in the first place. The simplified quantization and electrical noise model derived in eq. (2.3.5) does not hold well for the combination of the rising and flat part of the noise spectra (fig. 12). The slope of the f^{-1} electronic noise fits well with the data, however, a more suitable model is one with more degrees of freedom [*Sleeman*, 2006]. In any case, the IDC Hydroacoustic Global High and Low Noise Models [*Brown et al.*, 2012] provide a much tighter constraint and will generally always fall above any fitted quantization noise model, rendering the digitizer quantization noise constraint redundant.

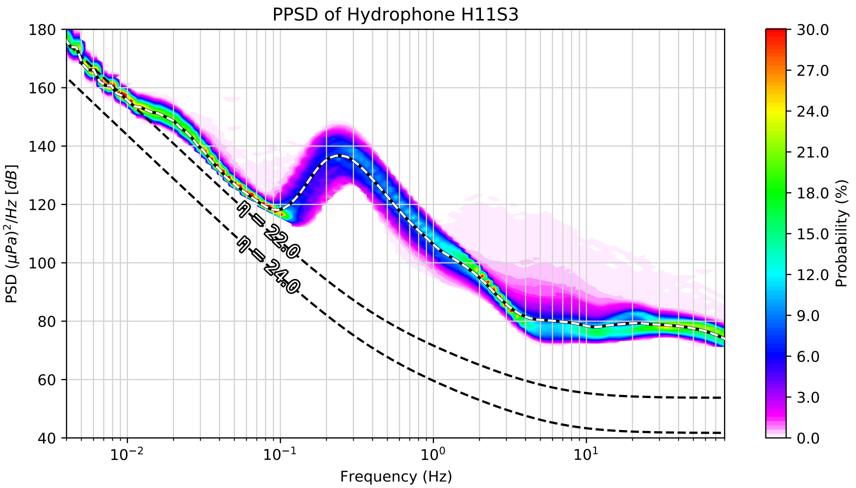

Figure 2.12: *PPSD of hydrophone H11S3 deployed in the network of International Miscellaneous Stations (IMS). The hydrophone is of type High-Tec HTI-90-U and processed with a segment length of* 1 h *using a reference of* $1\,\mu\mathrm{Pa}^2\,\mathrm{Hz}^{-1}$. *The instrument response is evaluated using the poles and zeros information from the metadata, using a maximum voltage of* 3.2768 V *as specified by the manufacturer of the data acquisition system.*

Tiltmeters

Tiltmeters are commonly used in monitoring infrastructures on the flanks of active volcanoes to detect periods of surface deformation caused by e.g., periods of inflation and deflation [*Dzurisin*, 2003]. The added value of observations from tiltmeters was explored in monitoring the Alkmaar gas field for subsurface deformation induced by pressure variations in the gas reservoir back in the end of the 20$^{\text{th}}$ century [*Sleeman et al.*, 2000]. The instruments are no longer in use in favor of a dense network of geophones and accelerometers. The working mechanism behind the tiltmeter is similar to the principle of an accelerometer, but instead of measuring ground motion it records the angle between the vertical component of gravity and the surface normal of the instrument. This measurement was historically done using a pendulum, but in modern tiltmeters the most accurate measurement is accomplished through an optical bubble level. Inside tiltmeters, the equivalence principle states that a tilt of θ radians provides an identical record as a horizontal acceleration of:

$$a(t) = \mathrm{g} \tan(\theta(t)) \tag{2.3.9}$$

where g is the local gravitational acceleration ($\approx 9.81\,\mathrm{m\,s^{-2}}$ in the continental Netherlands). Using eq. (2.3.9), the observed tilt in radians is converted to a virtual horizontal acceleration (fig. 2.13). The instrument response is flat to acceleration, and the maximum amplitude A is specified by the manufacturer datasheet at 330 µrad. A theoretical quantization noise model can be fitted using eq. (2.3.5) using a flat response to ground acceleration resulting in to $\eta = 17.0$, closely matching the specification of the manufacturer of the internal 16-bit datalogger. The microseisms are clearly observable in the instrument too and is also recommended to be used as a constraint on the median of the PPSD.

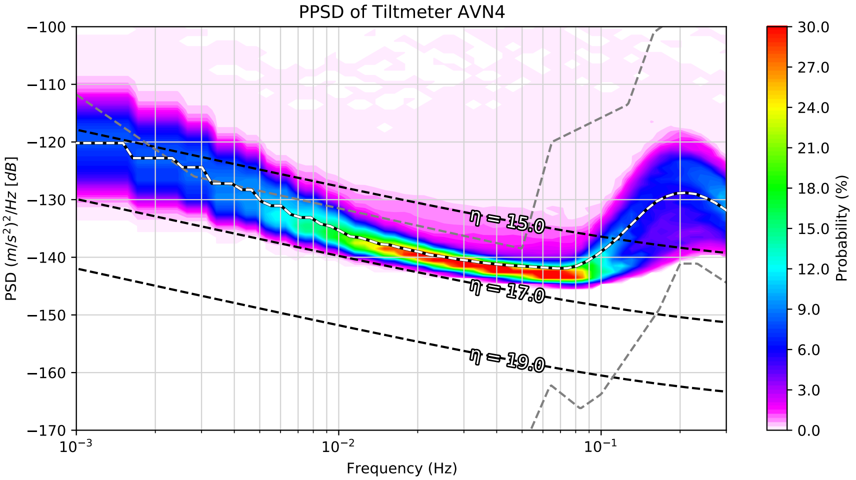

Figure 2.13: *PPSD of Applied Geomechanics LILY Tiltmeter (Serial Number: 8209) deployed in Oklahoma for 2019 using a segment length of* 1 h*. The Peterson [1993] noise models are illustrated in dashed grey. The number of fitted proxybits η comes out to approximately 17.0.*

Relative Gravimeters

At present, the NSAN does not employ gravimeters for geophysical monitoring in its operational infrastructure, but acknowledges the potential for e.g., volcano monitoring in the Caribbean Netherlands, similar to what is being explored in the NEWTON-g project [*Carbone et al.*, 2020; *Middlemiss et al.*, 2016; *Ménoret et al.*, 2018], or for hydrothermal monitoring purposes [*Sugihara and Ishido*, 2008].

The working principle behind the relative mechanical spring gravimeter is identical to that of an accelerometer, and superconducting gravimeters work by the levitation of a niobium sphere in a stable persistent magnetic field, creating a virtual non-mechanical mass-on-spring system [*Van Camp et al.*, 2017]. Compared to accelerometers, gravimeters are designed with a lower resonant frequency and are thus characterised by a much higher sensitivity [*Havskov and Alguacil*, 2016]. Its transfer function is flat to acceleration except the sensitivity drops proportional to ω^{-2} above the resonant frequency. The high sensitivity of the gravimeter at low seismic frequencies makes it easily saturated by e.g., surface waves from seismic events.

Compared to seismometers, for gravimeters it is less common to publish experimental transfer function estimates and generally a single flat sensitivity is used. This sensitivity is generally sufficient for the study of low frequency signals (e.g., earth normal modes, tides), which is the instrument's main frequency band of interest. In the seismic band, above the resonant frequency, it is necessary to correct data for the instrument frequency response when comparing to ground motion models. The recommendation of *Francis et al.* [2011] and others is emphasized, for operators to determine the full bandwidth frequency response of the instrument, similar to the *Network Of Superconducting Gravimeters* [1997], and publish the transfer function in widely used poles and zeros formats e.g., StationXML [*Ahern et al.*, 2015].

For the purpose of quality assessment, the data are processed using a segment length of 1 h, therewith eliminating frequencies below 0.001 Hz. The low frequency signals and spectral peaks (e.g., caused by earth tides) would be smoothed out regardless and contribute little value to quality control. The detection of changes in the seismic band is sufficient to confirm the instrument is performing as expected. The instrument in question (iGrav SG) clips at ± 10 V, with a sensitivity depending on configuration between 700 to 1000 $\mathrm{nm\,s^{-2}\,V^{-1}}$ below the resonant frequency, thus recording a maximum acceleration A of 7 to 10 $\mathrm{\mu m\,s^{-2}}$. Internally, a 24-bit ADC digitises the analog signal, placing the quantization noise far below the NLNM (fig. 2.14), also for frequencies below 0.001 Hz. The PPSD reaches the thermal noise floor due to Brownian motion in a mechanical oscillator at roughly −180 dB [*Warburton et al.*, 2010; *Rosat and Hinderer*, 2018], further suggesting that no quantization noise is recorded at any frequency. When corrected for the instrument frequency response, the microseisms are clearly visible and can serve as another constraint on the PSD estimate for gravimeters.

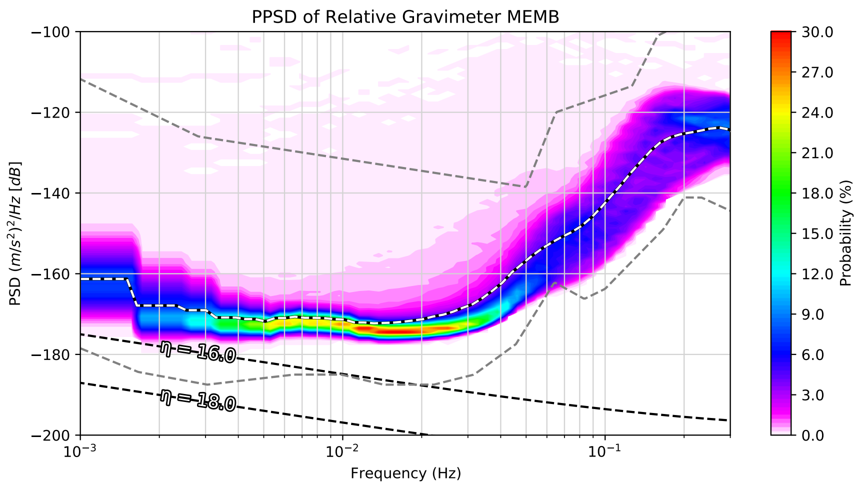

Figure 2.14: *PPSD of station MEMB (GWR C021 iGrav) superconducting gravimeter deployed in Membach, Belgium for 2019. The data has been corrected for the instrument acceleration frequency response over its full bandwidth. The white dashed line is the median of the PPSD. The quantization noise model from eq.* (2.3.5) *with two values for proxybits η is shown in black, where the internal ADC has 24 available bits. The Peterson [1993] noise models are illustrated in dashed grey.*

GNSS Receivers

The ground displacement observations from GNSS receivers in the NSAN are utilised for the detection of volcanic deformation on the islands of Saba and St. Eustatius in the Caribbean Netherlands. Ground displacement data can be derived from high temporal resolution (> 1 Hz) GNSS instruments that continuously record the position of the receiver against a reference earth ellipsoid. GNSS precise point position (PPP) solutions have inherently low precision, with a resolution on horizontal displacements between 2 mm to 4 mm and vertical displacements at the sub-centimeter level [*Xu et al.*, 2013]. Displacement solutions are often characterised by high noise levels due to e.g., the variable number of satellites used for the inversion, environmental multipath reflections, and atmospheric variations. Despite the inherently low precision, GNSS data may contribute to the detection of large seismic events or rapidly occurring volcanic phenomena, such as caldera collapse [*Elósegui et al.*, 2006; *Wang et al.*, 2013; *Neal et al.*, 2019]. For the purpose of long-term volcano monitoring, the receiver position is conventionally averaged out to a single position per day with a higher precision. Raw GNSS data were processed using the PPP algorithm, using the open source RTKLIB package [*Takasu*, 2013] to find vertical displacement ground motion with a 1 Hz sampling rate. To express the ground displacement PSD in acceleration the estimate is multiplied by ω^2 to differentiate from displacement to acceleration in the frequency domain where the resulting PPSD is illustrated in fig. 2.15.

Because of the large scatter in observations, quantization noise is not expected to be visible in the PSD estimate and no theoretical noise model can be fitted following eq. (2.3.5). Furthermore, no global noise models exist and for lack of better alternatives, the most effective remaining metric is the percentile criteria. Despite the proven contribution of this technique, automated GNSS anomaly detection is presently not included in the operational chain.

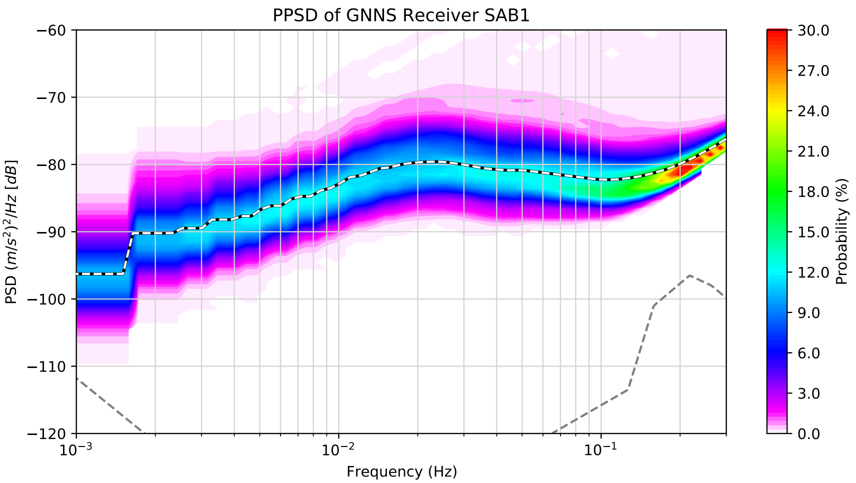

Figure 2.15: *Showing the PPSD of the vertical PPP displacement solutions (expressed in acceleration) of GNSS station SAB1 from 2019, deployed on the volcanic island of Saba in the Caribbean Netherlands. The white dashed line is the median of the PPSD. The Peterson [1993] NHNM is illustrated in grey.*

■ 2.3.3 NSAN Instrument Quality Verification Procedure

In the following section the quality control procedure and constraints that are applied to specific sets of instruments are discussed. The procedure is operated weekly on the latest monthly PPSD for each instrument. Not all instruments discussed in section 2.3.2 are currently operational in the NSAN and this section is limited to the types of instruments that are. A flowchart of all the imposed constraints from section 2.3.1 on the operational network is schematically visualised in fig. 2.16. This process can easily be extended for other types of instruments.

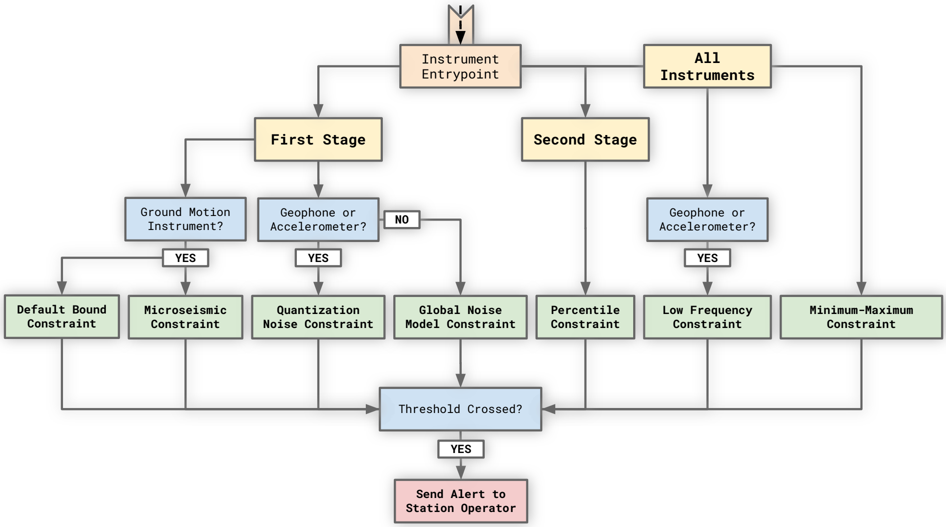

Figure 2.16: *Schematic flowchart illustrating the PPSD quality assessment algorithm. The process is divided into two stages based on the availability of verified archived data and thus percentile confidence intervals. The algorithm can easily be extended for other types of data and the recommended quality criteria discussed in sections 2.3.1 and 2.3.2*

.

All Instruments

For all instruments it is confirmed there is a minimum scatter of 5 dB at $\sim$ 3 Hz. If the instrument is an accelerometer or geophone the system verifies that the low-frequency constraint is passed. The processing is then split into different paths for instruments in the different stages of quality control.

Instrument Stage Zero: New Installations

Instruments in the first month of operation are placed in this stage and do not contribute to operational workflows. This stage exists to evaluate the performance of an instrument in its environment before its data are used and published. This stage shares the metrics defined in stage one.

Instrument Stage One: Recent Installations

Instruments with less than one year of archived quality control passed data are kept in the first stage. Because archived data are absent for new stations the system falls back to comparison against generic limits following the flowchart illustrated in fig. 2.16. For stations in the first stage, the median of the PPSD of ground motion instruments is compared against the expected microseisms at 0.3 Hz, and the default upper bound over all frequencies. The theoretical quantization noise constraint is applied for accelerometers and geophones where a theoretical lower bound of the instrumental noise is available. Instead, for broadband seismometers and infrasound stations the system compares the median against the instruments respective global noise model. Instruments in stage one are promoted to stage two if more than a year of high quality data has been archived.

Instrument Stage Two: Existing Installations

Stations in the second stage have their latest monthly median of power spectra compared against archived data that includes constraints on the local noise field around the instrument. The PPSD of validated archived data provides an upper and lower 2.5 % confidence limit on the incoming data.

2.4 Results

2.4.1 Identification of Degraded Instrument Performance

In the following section three examples of degraded instrument performance are presented that were identified by the system in an operational setting.

Detection of Pure Instrumental Noise Recordings

Instruments with shorted inputs or dataloggers that are disconnected from the sensor only record intrinsic electrical noise. The power spectral density is expected to be stable and narrow over the full bandwidth. Geophone FDG1 shows an example of an instrument identified by the system that is recording only digitizer noise (fig. 2.17). This example fails the required minimum-maximum difference of 5 dB threshold.

High Noise at Low Frequencies

The PPSD of surface accelerometer BLOP illustrated in fig. 2.18 should be dominated by instrumental noise at lower frequencies. However, a clear scatter can be identified and the PPSD is nowhere near the expected quantization noise lower limit. Another undesirable process is introducing high levels of ambient noise and the cause should be investigated.

Percentile Threshold Trigger

Changes in the trend of the PSD estimates of an instrument can be detected by comparing statistical parameters of the latest PPSD against confidence percentiles calculated from previously archived spectra. The PPSD of station DR023 in fig. 2.19 shows the mean of the distribution of the latest month of data falling outside the expected percentile. This divergence of the ambient noise level was automatically detected and the station operator was notified. After this detection, the percentile constraint was changed to use the median of the PPSD instead of the mean.

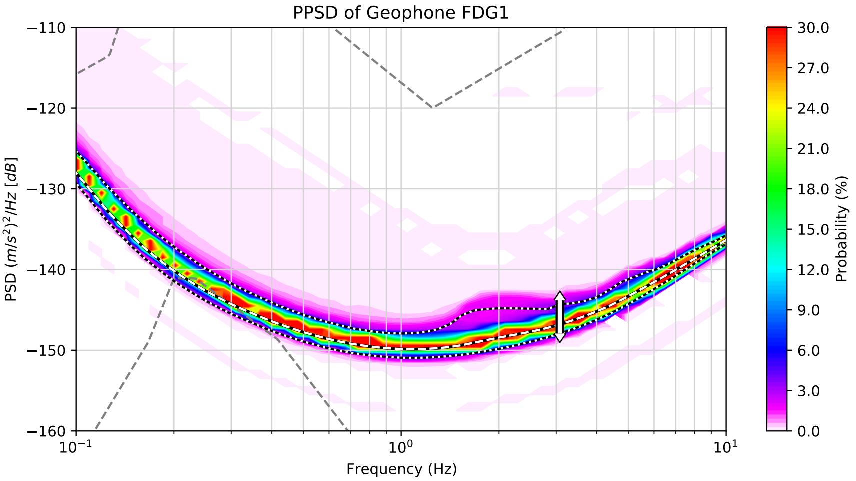

Figure 2.17: *Showing geophone FDG1 (N-component) throughout 2019 that is dominated by instrumental noise over its full bandwidth. The range of the white arrow at* $\sim 3\,\mathrm{Hz}$ *covers a* $4\,\mathrm{dB}$ *difference between the minimum and maximum that is insufficient to pass the minimum-maximum threshold of* $5\,\mathrm{dB}$ *and was raised for review by the system.*

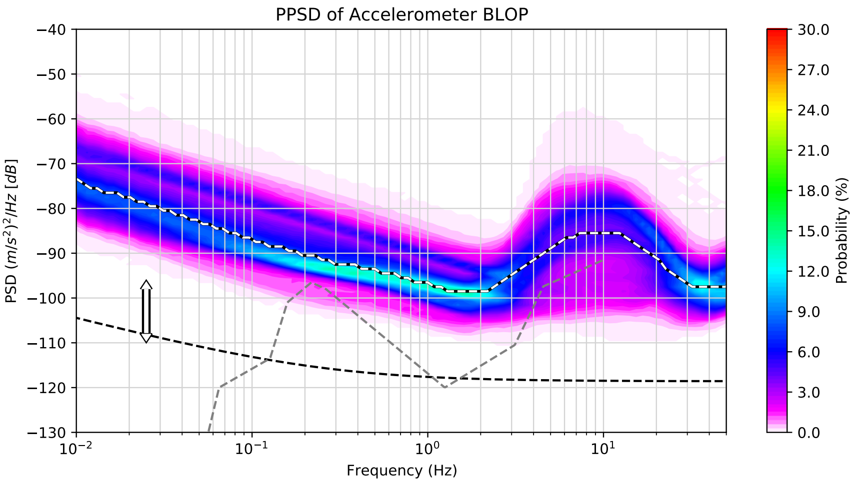

Figure 2.18: *The median of the PPSD (2019) of surface accelerometer BLOP (vertical) does not fall within the required low frequency threshold. The median of the distribution exceeds the theoretical model from eq.* (2.3.5) *(dashed black line) for a batch-1 accelerometer (table 2.1) by more than* $10\,\mathrm{dB}$ *(white arrow) at* $0.025\,\mathrm{Hz}$ *and was therefore raised by the system.*

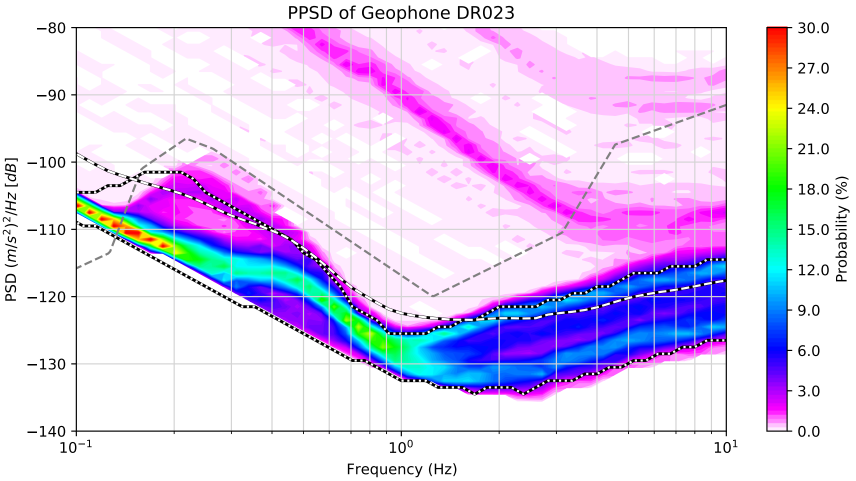

Figure 2.19: *The mean of the PPSD (dashed) of geophone DR023 (horizontal) during September 2020 exceeds the maximum* (97.5^{th}) *archived percentiles (dotted) around* 1 Hz*, due to anomalously high PSD estimates during the latest month. The percentiles were calculated from validated archived data.*

2.5 Discussion

The amount of data flowing into the archive is growing rapidly, and network operators must increasingly rely on automated and unsupervised processes to detect and identify instrumental anomalies. The main objective of the PSD database is to store spectral estimates as input for the automated quality control process for the NSAN. The spectral smoothing and binning makes the database less suitable for anomaly detection before the higher seismic frequency band (< 0.01 Hz), and for scientific purposes that require accurate resolution of spectral peaks [*Anthony et al.*, 2020]. Nonetheless, the system is designed as a data product for researchers and analysts that provides conventional PSD estimates for the entire NSAN archive without requiring any client-side resources for processing. The API is designed to aggregate and visualize PSD segments in various ways (e.g., PPSD, spectrogram, power time series at a particular frequency) depending on the user preference. The presented database of individual PSD estimates is the most versatile and flexible way to cache spectral information. The fine granularity of the hourly segments makes the database valuable for studying transient signals present over a sufficiently large bandwidth, for example looking at the footprint of anthropgenic noise [*Lecocq et al.*, 2020], or developing and extending global noise models [*Wolin and McNamara*, 2019].

Because the PSD estimates are cached in the database, no processing is required except for the retrieval of the segments. One problem this approach introduces is that the instrument frequency response is already deconvolved from the data before it is stored. When the instrument transfer function is corrected by the operator, all cached PSD segments for this channel become invalid and need to be scheduled for reprocessing. Another option to consider is to store the PSD in counts and removing the instrument response during the request, adding some additional overhead but eliminating the need for reprocessing when metadata is altered.

Because the procedure is run weekly, using the past month of data, a delay in the detection is expected from the moment an instrument begins to perform poorly. The monthly median of the PPSD is used so that the PSD segments are averaged out for short-lived transient effects that may trigger the system as a false positive. An alternative consideration is to run the system at a daily interval. However, because of the slow evolution of the median of the PPSD, this will not significantly advance the moment of detection. Furthermore, with over 700 stations, having to review the results of the system trigger every day exceeds the available human resources at our disposal. Fine tuning of the system detection sensitivity is easily performed by changing configuration during the PSD calculation, e.g., segment length, smoothing range, fixed frequency interval. The sensitivity of the metrics can also be easily adjusted, by either raising or decreasing the thresholds for detection.

The presented database is highly efficient in terms of storage, amounting to only a total of less than 0.05% percent of the size of the NSAN data archive. In comparison, a database of uncompressed PSDs would occupy the same storage size as the waveform data does in the time domain. A significant reduction in size is introduced by the full-octave smoothing, fixed one-eighth frequency intervals, and rounding of power densities to the nearest dB to fit within a single-byte range. The presented custom compression scheme saves only a single 8-bit array per spectrum and the metadata required for the PSD reconstruction.

The number of proxybits η used to model the quantization noise of the datalogger is empirically derived from field data, using a part of frequency band that is dominated by instrumental noise. The simple theoretical models are based on purely white quantization noise superimposed on a pink noise spectrum (f^{-1}) that works well for accelerometers, geophones, and tiltmeters that are dominated by instrumental noise. More accurate models for instrumentation self-noise could be found by conducting laboratory experiments through e.g., shorting of the datalogger input and measuring its output. The presented approach is based on simple theoretical models [*Bennett*, 1948; *Sleeman*, 2006] and empirical fitting because the NSAN employs many different types of instruments of varying generations that are currently operational in the field. This constraint will be able to detect a mismatch between the sensor output voltage and the datalogger input voltage (table 2.1), or when a wrong instrument sensitivity is used. It should be emphasized that the system is unable to detect problems if the same error is made in computing $PSD_{\min}$ and the instrument metadata, and suggest the model fitting is done independently from the metadata entry. Datalogger quantization noise models could not be empirically fitted for broadband seismometers, gravimeters, infrasonic stations, hydrophones and GNSS receivers. For these instruments, in the frequency band the system applies the automated quality control, environmental or sensor noise dominates over instrumental noise.

The derived confidence range of -90 to $-140\,\mathrm{dB}$ for the microseisms is similar to the *Peterson* [1993] noise models at 0.3 Hz and may be partially redundant for the Netherlands. A more accurate model would include the relationship between coastal distances and the value of PSD estimate as illustrated in fig. 2.6, with a slope of approximately $-0.15\,\mathrm{dB\,km^{-1}}$. One downside of this change would be the detection of more false positives as the relationship is not fully consistent between all instrumental sites and does not include local site effects.

A number of different constraints for the detection of abnormal PSD estimates are applied, so that if one constraint fails to trigger as expected, the anomaly may be picked up by another independent metric that is based on a completely different characteristic of the PSD. In particular with the availability of confidence intervals per station calculated on archived data that has been verified by the station operator, the system will be able to identify deviations from ordinary conditions within a week. The system is designed to support extensions with other types of instruments e.g., hydrophones, tiltmeters, gravimeters, and GNSS receivers that are presently not included in the NSAN, but may be in the future.

2.6 Conclusion

This study presents an operational implementation of a quality verification procedure for the NSAN based on the automated analysis of PSD estimates. The system is designed to efficiently store PSD estimates in a database using a custom compression scheme. The NSAN is continuously expanding and interest in new and additional instrumentation is rapidly growing, thus highlighting the need for automated policies and procedures. A universal method is proposed to automatically verify the performance of many types of geophysical instruments in a technically similar way. The variation of PSD estimates through time from geophysical instruments serves as an effective mechanism to assess its performance. The commonly used technique of using PSD estimates for quality control is applied and extended where additional quality criteria of the PSD are defined and recommended for different instruments. These criteria are based on 1) conventional global noise models, 2) instrument specific models based on digitizer quantization noise, 3) regional models for the Netherlands using the microseisms, and 4) site-local using data-driven statistical confidence limits. For the NSAN the automated procedure is scheduled weekly and verifies that the latest monthly archived waveform data falls within the limits imposed by our quality constraints defined in section 2.3.1. This system proves promising for many geophysical instruments and can easily be adapted and extended in the future. It is shown that the system is able to monitor that instruments in the NSAN are operating as expected, and automatically detect degraded instrument performance at a national network scale.

2.7 Acknowledgements

We thank the R&DSA software development team Jarek Bienkowski, Gert-Jan van den Hazel, and João Paulo Pereira Zanetti for their assistance in implementing the software stack. Jelle Assink provided his advice to process infrasound and hydroacoustic data. Elske de Zeeuw-Van Dalfsen provided the GNSS data and valuable support. Hans van der Marel is thanked for his advice in avoiding pitfalls encountered during GNSS PPP processing. Daniele Carbone and Michel Van Camp are thanked for sharing their experience with the iGrav superconducting gravimeters. We would like to thank the editor and three anonymous reviewers for their comments and suggestions that have helped us improve the final manuscript. We express our sincerest gratitude to the late Dr. Peter Fox.

3

Passive Assessment of Geophysical Instruments Performance using Electrical Network Frequency Analysis

Abstract

The electrical network frequency (ENF) of the alternating current operated on the power grid is a well-known source of noise in digital recordings. The noise (i.e., signal) is widespread and appears not just in close proximity to high-voltage power lines, but also in instruments simply connected to the mains powers grid. This omnipresent, anthropogenic signal is generally perceived as a nuisance in the processing of geophysical data. Research has therefore been mainly focused on its elimination from data, while its benefits have gone largely unexplored. It is shown that mHz fluctuations in the nominal ENF (50/60 Hz) induced by variations in power usage can be accurately extracted from geophysical data. This information represents a persistent time-calibration signal that is coherent between instruments over national scales. Cross-correlation of reliable reference ENF data published by electrical grid operators with estimated ENF data from geophysical recordings allows timing errors to be resolved at the 1 s level. Furthermore, it is shown that a polarization analysis of particle motion at the ENF may assist in the detection of instrument orientation anomalies. While the source of the ENF signal in geophysical data appears instrument and site specific, its general utility in the detection of timing and orientation anomalies is presented.

Published as Koymans, M., Assink, J., de Zeeuw-van Dalfsen, E., Evers, L. (2023). Passive Assessment of Geophysical Instruments Performance using Electrical Network Frequency Analysis. Seismica, 2(2), doi: 10.26443/seismica.v2i2.1082.

3.1 Introduction

Sustaining reliable and continuous operation of instruments in the field is a key objective in the maintenance of geophysical monitoring infrastructures. This objective is particularly challenging for remote deployments, and equipment that cannot easily be accessed, e.g., for sensors buried at depth inside seismic boreholes. Active assessments that involve station maintenance visits are costly, time-consuming, and require perpetual planning and effort. Methods for passive quality assessment are often pursued due to their advantages in terms of scalability and reduced cost [*McNamara and Boaz*, 2006a; *Ahern et al.*, 2015; *Ringler et al.*, 2015; *Trani et al.*, 2017; *Petersen et al.*, 2019; *Pedersen et al.*, 2020; *Koymans et al.*, 2021]. Moreover, such passive techniques do not disturb the measurement setup itself and may be useful in e.g., citizen science [*Raspberry Shake, S.A.*, 2016] where the acquisition of high quality data can not be guaranteed. In the case where correction factors can be estimated, they can also be retroactively applied to an archived dataset. Data assessment is not exclusively useful to science, but also serves a purpose to detect malicious actors and data tampering that is critical in e.g., the verification of the Comprehensive Nuclear-Test-Ban Treaty [*Coyne et al.*, 2012].

Geophysical data may express characteristic spectral peaks that emerge from the electrical network frequency (ENF) of the alternating current (AC) operated on the electrical grid. This signal is sometimes referred to as *powerline* noise, but notably does not appear exclusively near high voltage power lines and is widespread. The signal is omnipresent in recordings from e.g., seismometers [*Bormann and Wielandt*, 2013], gravimeters [*Imanishi et al.*, 2022; *Křen et al.*, 2021], microbarometers, and other digital instruments that are connected to or deployed nearby any type of electrical infrastructure or mains power supply. The ENF signal is usually perceived as a nuisance during the processing of geophysical data, and research has mainly been targeting its elimination [*Butler and Russell*, 1993; *Xia and Miller*, 2000; *Levkov et al.*, 2005]. For most purposes, the application of a narrow band-stop (notch) filter is sufficient to remove the signal. However, ringing artefacts, higher harmonics, or overlap with the bandwidth of interest sometimes makes the application of such filters impractical. For example, in seismoelectric acquisition and seismic exploration, advanced methods for the removal of coherent electrical noise are applied [*Butler and Russell*, 1993, 2003]. While methods to eliminate the ENF signal from geophysical data are well known, the benefits of its presence are rarely explored. This study approaches the ENF from a different perspective, and demonstrates its utility as a signal in geophysics.

In this manuscript, two benefits of detecting the ENF in geophysical data are explored and used as passive quality assessment tools. First, the background information on the ENF is described (section 3.2), followed by an introduction of the data sets that are used (section 3.3.1). After that, the methodologies are described to **(i)** extract the ENF signal from spectrograms of geophysical data (section 3.3.2) and compute cross correlations (section 3.3.3), and **(ii)** complete a polarization analysis of the particle motion around the ENF (section 3.3.4). Results from cross correlations between spectrogram-estimated and reference ENF data are presented, demonstrating that timing errors with a resolution near the 1 s level can be resolved and verified (section 3.4.1). The accuracy of the recovered timing discrepancies are statistically quantified (section 3.4.2) and checked using teleseismic arrivals (section 3.4.2) that should be observed simultaneously on stations in close proximity, providing an alternative way of detecting relative time shifts. Results from the polarization analysis indicate that the method is capable of detecting gross sensor orientation anomalies (section 3.4.3). Finally, the source of the ENF signal in different geophysical instruments is discussed, and comments are provided on possible future avenues of research (section 3.5).

3.2 Background

3.2.1 Electrical Network Frequency

An abridged description of the electrical grid concerns power generators that supply electrical energy to consumers. Conceptually, generators are rotating turbines with magnetic cores that induce AC in coils following Faraday's law of electromagnetic induction. All generators on the grid collectively produce synchronous AC, with waveforms that are equal in amplitude, phase, and frequency. Because the electrical energy produced by the generators cannot be stored it must be immediately consumed, requiring a delicate balance between production and demand. At an instant when more energy is consumed than produced, the required excess power is drawn from the rotational inertia of the generators. This synchronously reduces the rotation speed of the generators on the grid, and subsequently lowers the effective ENF. Likewise, a sudden decrease in load causes the turbines to spin faster, leading to an increase of the ENF. Electrical grid operators balance the amount of electrical work done by the generators with the demand of consumers to keep the ENF stable at 50 Hz for continental Europe and 60 Hz for the United States. This balance is diligently maintained, and operational procedures are in place to limit deviations from the target ENF to within 10 to 50 mHz.

All electrical components – including geophysical instruments – are to some degree susceptible to the secondary effects of the AC operated on the electrical grid (fig. 3.1). Signals may be incurred from stray electromagnetic fields that are emitted from nearby current carrying wires and operating electronics. Common sources of the ENF signal being carried over in electric devices are through ground loops, and by direct electromagnetic induction of poorly shielded wires and circuitry. Magnetostriction in transformers [*Gange*, 2011] and full-bridge rectifiers (AC → DC) in power supplies may produce vibrations and audible sound at double the ENF. The well-known audible sound originating from the ENF is commonly referred to as *mains hum*. In broadband seismometers, a known coupling mechanism is through the suspension spring that responds to changing magnetic fields [*Forbriger*, 2007]. Intense changing magnetic fields may even cause the housing of instruments to vibrate [*Klun et al.*, 2019]. At frequencies above the operated ENF, overtones at integer multiples of the ENF can sometimes be observed [*Cohen et al.*, 2010; *Schippkus et al.*, 2020].

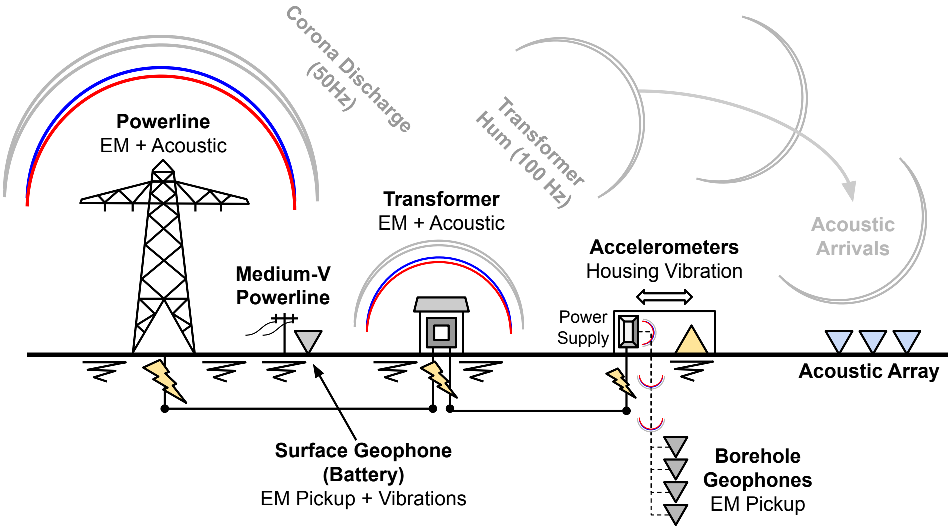

Figure 3.1: *Overview of suspected sources of the ENF signal in geophysical data where the colors represent electromagnetic (red/blue), acoustic (grey), and seismic (black) coupling. The coupling mechanism varies between instruments and installation site. The signal may be coupled through physical vibrations, acoustic waves, or by direct magnetic induction.*

While the ENF signal is typically of minor influence, equipment that integrates amplifiers may boost it to significant amplitudes. While the source of the ENF signal in high gain equipment is not always directly apparent from its surrounding, its persistence and omnipresence remains remarkable.

3.2.2 ENF Analysis

ENF analysis typically concerns the detection of mHz variations of the ENF in digital recordings as a function of time, of which an example is illustrated in fig. 3.2. These variations can be extracted from, e.g., audio [*Cooper*, 2008], optical [*Garg et al.*, 2011], and geophysical data [*Cohen et al.*, 2010]. Because the AC is operated synchronously and uniformly on the electrical grid, digital recordings of the ENF represent a fingerprint that is coherent nationwide, and because of effectively random load fluctuations, represents a signal that is unique in time. The estimated variations in the ENF from digital recordings may thus be compared to an independent reliable reference measurement of the ENF that is provided by electrical grid operators. Such analysis of the ENF has been used to timestamp audio recordings [*Garg et al.*, 2012] and confirm the authenticity of digital records. The successful use of ENF analysis as forensic evidence [*Cooper*, 2010] is a testament to the effectiveness and reliability of the technique.

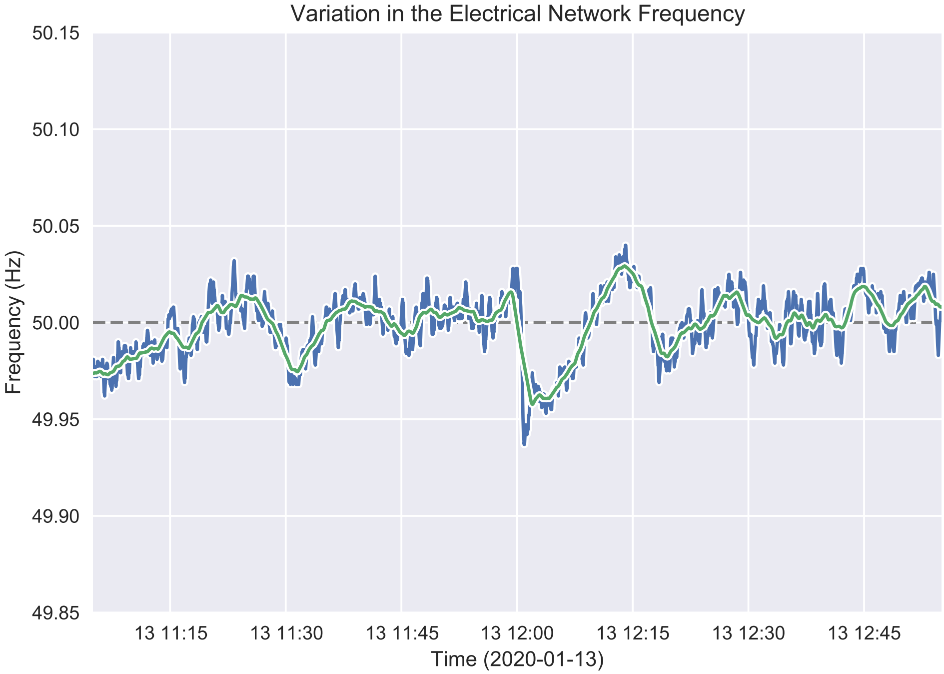

Figure 3.2: *Example of minor* mHz *variations in the ENF during two minutes on Jan 13th, 2020 around the nominal European grid frequency of* 50 Hz *(grey dashed line). These data were not recorded by a geophysical instrument but illustrate reference ENF data that were downloaded from electrical grid operator TransnetBW. The raw data are plotted in blue, with a smoothed* 150 s *moving average illustrated in green.*

3.3 Methodology

3.3.1 Instruments and Data Used

Various data types from different sensors are analysed in order to study the specific character of the ENF in these instruments. Data from the Netherlands Seismic and Acoustic Network [*KNMI*, 1993] and E-TEST temporary deployment [*Shahar Shani-Kadmiel et al.*, 2020] (fig. 3.3 and table 3.1) are treated. The G-network of the Netherlands Seismic and Acoustic Network (NSAN) consists of nearly seventy 200 m deep boreholes in the Groningen province with geophones installed at 50 m depth intervals, and an accelerometer located at the surface. Data from the NSAN that belong to a low-frequency acoustic array installed at the Royal Netherlands Meteorological Institute in De Bilt (see supplementary information), and seismo-acoustic arrays at LOFAR sites in Drenthe are also analysed. The E-TEST temporary deployment consists of a dense array of battery-operated surface geophones located in the south of the province of Limburg without a mains power supply.

Table 3.1: *Descriptions and characteristics of geophysical instruments used for various aspects of ENF analysis that are treated in this manuscript. Instrument and response details are accessible from FDSN webservices (https://rdsa.knmi.nl and https://orfeus-eu.org).*

Sensor and Description	**Network**	f_s
SM6H Borehole Geophone	G-network (NL)	200
Kinemetrics EpiSensor Accelerometer (ES-T)	G-network (NL)	200
Sensor B.V. SM-6/U-B 4.5Hz 375 Geophone	LOFAR Array (NL)	250
Hyperion low-frequency sound microphone	De Bilt Array (NL)	500
SENSOR Nederland, PE-6/B, 3C battery geophone	E-TEST (3T)	500

3.3.2 Spectrogram Calculation and ENF Estimation

Independent ENF reference measurements at 1 Hz are universally accessible and downloaded from, e.g., the power-grid frequency database [*Gorjão et al.*, 2020] and the website of TransnetBW GmbH. In this manuscript, ENF measurements from a German provider were used – data that are synchronous with the electrical grid operated in the Netherlands. The reference ENF data were smoothed using a centered moving average filter over 150 s (e.g., see fig. 3.2).

Geophysical data from the instruments summarised in table 3.1 were pre-processed using ObsPy [*Beyreuther et al.*, 2010] (read and merged) and spectrograms were calculated using the SciPy *spectrogram* method [*Virtanen et al.*, 2020] with a segment length of 150 s, employing a 50 % overlap between consecutive segments. It was determined empirically that this segment length provided the most effective trade-off in resolution between time and frequency to resolve the ENF from the spectrograms. A linear trend was removed from each segment and the data were tapered using a cosine window with a shape parameter of 0.25. A Gaussian filter was applied in the frequency domain before the ENF was estimated from the spectrogram. This filter represents the mean and standard deviation of the yearly ENF signal ($\mathcal{N}_{50}(\mu, \sigma) =$ 50.000 Hz, 441×10^{-4} Hz), and eliminates peaks in the spectrogram that are likely unrelated to the ENF. For each segment, the estimated ENF is represented by the frequency bin that associates with the maximum PSD within the 49.85 to 50.15 Hz band. An identical approach (with modified filter $\mathcal{N}_f$) was used for the extraction of overtones of the ENF in higher frequency bands (e.g., at 100 Hz).

3.3.3 Cross-Correlation Analysis

The estimated variations in the ENF were interpolated to 1 s and cross-correlated with independent reference ENF data. A negative delay from the cross-correlation result implies that the reference signal leads the estimated ENF and is therefore behind *true* time. A statistical analysis of the accuracy and precision of the method was completed using an ensemble of cross correlations from instruments that are known to have zero time delay. The accuracy of the method and the recovered timing errors were further verified at a seismic array using teleseismic arrivals from

Figure 3.3: *Map of the Netherlands showing four groups and locations of geophysical instruments in the field (G-Network – geophones and surface accelerometers; E-TEST Deployment – battery operated geophone nodes; LOFAR – seismo-acoustic array; De Bilt – acoustic array). Further details on the instruments are provided in table 3.1*

an event near the Kermadec Islands, New Zealand (2021-03-04T19:28:33 UTC).

Because the teleseismic arrivals are characterised by a near vertical incidence angle, the arrival times for proximal stations are expected to be similar, providing an alternative relative timing reference to compare against the obtained ENF analysis results.

3.3.4 ENF Polarization Analysis

Another independent aspect where the ENF signal can be leveraged is for surface accelerometers in the G-network that express a significant and strongly polarized susceptibility to the ENF. Accelerometer data were rotated towards a north-east orientation following the azimuth provided by the station metadata. The polarized ENF signal was isolated with a zero-phase band-pass filter between 49.85 to 50.15 Hz. A principal component analysis (PCA) was applied to the three-dimensional particle motion data and eigenvalues (λ_1, λ_2, λ_3) were recovered, from which the degree of rectilinearity [*Jurkevics*, 1988] was calculated:

$$1 - \left(\frac{\lambda_2 + \lambda_3}{2\lambda_1} \right) \tag{3.3.1}$$

The azimuth of the principal direction of motion (θ) was derived from the largest eigenvector $\mathbf{u}_1$, as given by its north and east components: $\theta = (\mathbf{u}_{1_\mathrm{N}}, \mathbf{u}_{1_\mathrm{E}})$. The goal of this method is to investigate whether the ENF can be used to verify the instrument orientation as specified in the station metadata.

3.4 Results

3.4.1 Timing Errors from ENF Analysis

An example ENF analysis for instrument EpiSensor accelerometer G180 is shown in fig. 3.4. The figure illustrates the reference variation in the ENF around 50 Hz (A), the raw seismometer spectrogram expressed in ground acceleration (B), the spectrogram with the Guassian filter applied (C), and that the ENF can be accurately recovered from the filtered spectrogram (D) and compared to the measured ENF (E). fig. 3.5 panel A shows the measured and estimated ENF time series from fig. 3.4. The curves were vertically displaced from an average of 50 Hz to illustrate their similarity. The full cross-correlation of the measured and estimated ENF is illustrated in panel B and expresses a peak at a delay of −1 s (C), meaning the instrument effectively runs behind *true* time. An identical analysis for an acoustic station is presented in the supplementary information because of additional complications that were encountered.

The presented example result in figs. 3.4 and 3.5 illustrates the method for a single instrument, but the approach has been successfully applied to all instruments in the NSAN network, including surface accelerometers, geophones, and microbarometers. The results indicate that the proposed method appears capable of detecting misfits between the estimated and reference ENF in geophysical data, potentially providing a stable nationwide timing calibration signal.

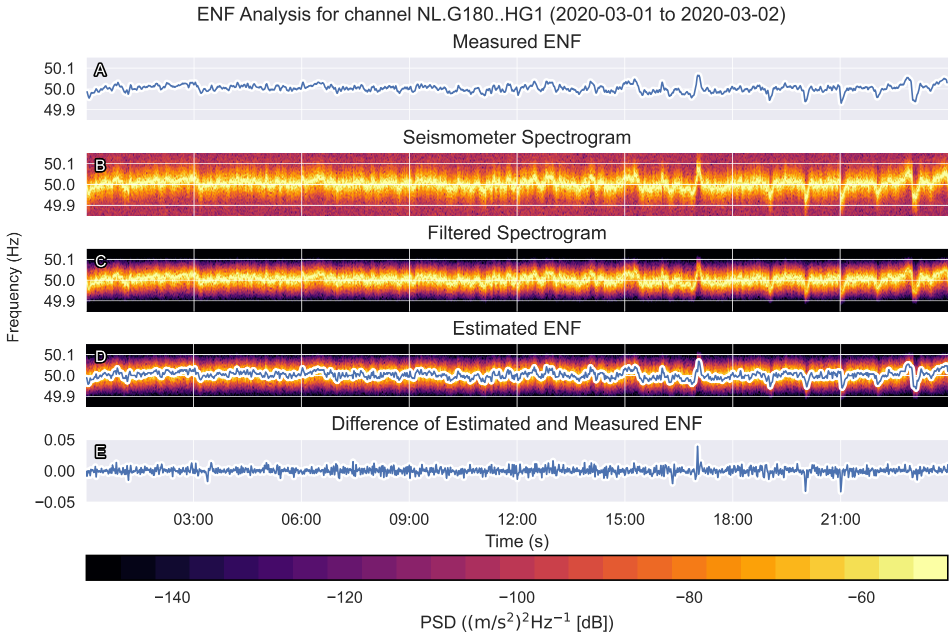

Figure 3.4: *A) The reference ENF downloaded from the website of TransnetBW GmbH. B) Acceleration spectrogram of EpiSensor NL.G180..HG1 (Groningen, the Netherlands) between 2020-03-01 and 2020-03-02. C) The modified spectrogram using a simple Gaussian filter. D) The estimated ENF from the filtered spectrogram derived from the maximum PSD of each time segment. E) The difference between the estimated and measured ENF.*

3.4.2 Confirmation of Timing Error Results

In the following sections, two methods are used to assess the precision and accuracy of the proposed method for the detection of timing anomalies.

Resolution of the Method

An estimate of the statistical significance of the recovered time lags is obtained through an ensemble of cross correlations between the measured and estimated ENF from all components of 71 surface accelerometers in the NSAN. These instruments are known to have accurate timestamps because they obtain timing through GPS and should thus express a zero-second delay from true time. Figure 3.6 shows an ensemble of 211 cross correlations with its average and 95 % confidence interval in blue. The peaks of all cross correlations and recovered time lags are also illustrated by grey markers. Accelerometers for which the ENF could not be resolved due to poor data quality or elevated noise have been removed from the ensemble. The majority of instruments express a lag of −1 s between the estimated and measured ENF data, while the others express a 0 s time lag as expected. The confidence

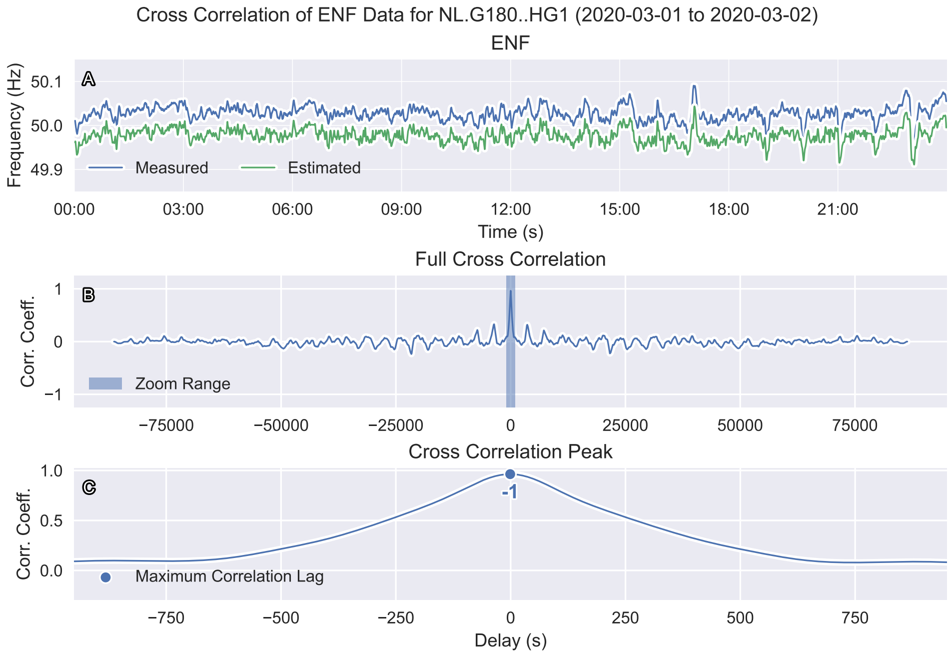

Figure 3.5: *A) Comparison between the reference ENF (blue) provided by TransnetBW GmbH and the estimated ENF (green). Note that the data have been offset from the mean of* 50 Hz *for illustrative purposes to show their similarity. B) The full cross correlation between the estimated and reference ENF. C) Zoom in on the blue span around the correlation maximum (≈ 0.96) with the recovered peak and time delay indicated (*−1 s*).*

interval on the mean time lag from this ensemble illustrates the estimated accuracy and precision of the method at approximately 1 s. Furthermore, the repeatability of the methodology between 211 data channels is a testament to its consistency. The minor stable deviation from the expected delay of zero may be caused by a non-precise or rounded off timestamp of the ENF data provided by the grid operator.

Confirmation of Results Using Teleseismic Arrivals

The accuracy of the recovered timing errors was further verified using teleseismic arrivals at geophone ENV1 and nearby LOFAR arrays L106 and L208 of the M8.1 earthquake near the Kermadec Islands, New Zealand that occurred at 2021-03-04T19:28:33 UTC. The first two rows of fig. 3.7 show station ENV1 and L2082 at 24 km and 13 km distance from LOFAR array L106 (bottom 6 rows) respectively. The predicted seismic arrival times for the PKIKP phase of the event were calculated with TauPy [*Beyreuther et al.*, 2010] using the IASP91 model [*Kennett and Engdahl*, 1991]. The left column in fig. 3.7 shows that the recorded arrivals of the seismic phase of the original time-series are misaligned. The right panels show the same data

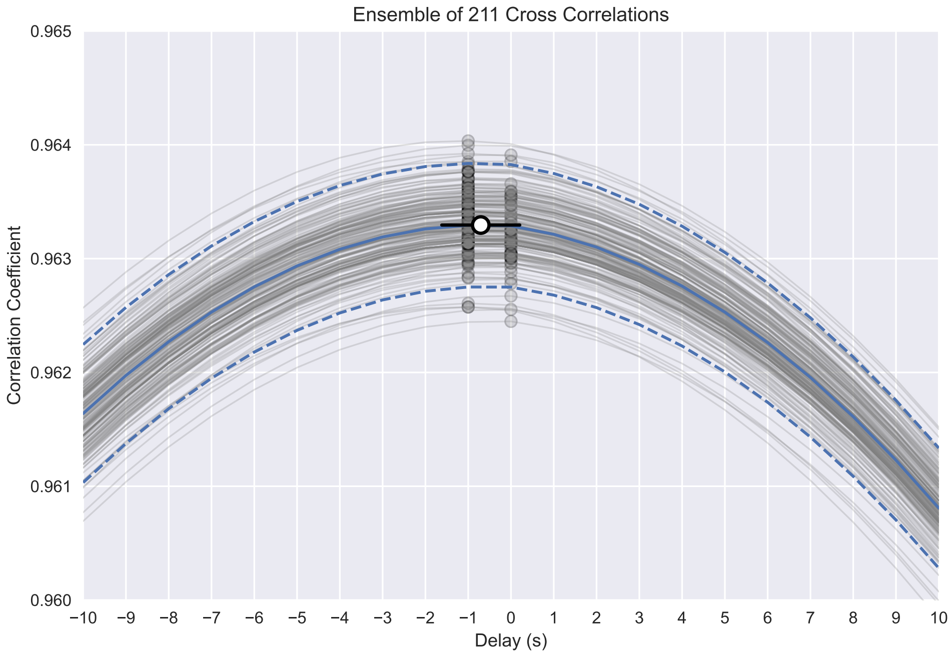

Figure 3.6: *Average and 95 % confidence limits (blue curves) of an ensemble of 211 cross correlations with the measured ENF (grey curves) for all components of all Groningen surface accelerometer in the G-network on 2020-03-01. The accelerometer data have accurate timestamps and should resolve to a zero-time delay. The grey markers indicate the recovered peaks from the cross correlations and hence the respective delay times with the measured ENF. The black marker represents the mean time lag and 95 % confidence interval, illustrating the accuracy and precision of the method approaches* 1 s.

shifted by the recovered delay from the ENF analysis (marked in the top-left corner of each panel). Geophone ENV1 and LOFAR station L2081 acquire timing through GPS and have near zero delay, while the L106 geophones express between −21 to −7 s delays with the reference ENF. This effect is unsurprising as the instruments use the Network Time Protocol (NTP) instead of GPS and may experience clock drift over time without a stable internet connection. With the expected timing corrections applied, the alignment of the arrivals is vastly improved. The remaining misalignment may be a consequence of local geology and site-response, and the inherent 1 s resolution limit of the technique. Furthermore, the timing misfits from the ENF analysis were calculated over 24 h while the timing error of the L106 array was observed to vary by multiple seconds in a day. At the time of the teleseismic arrival, the ENF delay appeared to be consistently 6 s behind the reference data for the entire NSAN network. This effect was corrected in fig. 3.7 using an average of many GPS locked stations.

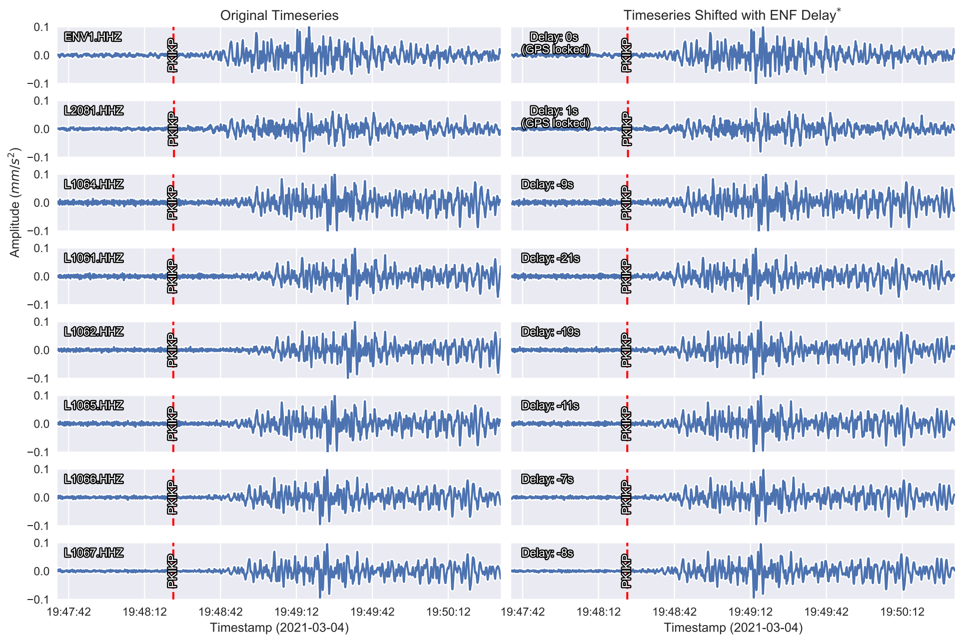

Figure 3.7: *Comparison of GPS locked station with near zero delay (ENV1, L2081) with local array L106 (bottom 6 rows) recording a teleseismic arrival. The expected arrival times for PKIKP phase is illustrated. The right columns shows the same traces shifted by the recovered timing error from the ENF analysis. *Note: all delays were corrected for a consistent* $-6\,\mathrm{s}$ *offset across the entire network that appeared present on that day.*

3.4.3 Orientation Errors from ENF Analysis

A polarization analysis of the ENF signal was applied to three-component data from surface accelerometers in the G-network. A principal component analysis provides the dominant modes of variance of these data (i.e., the dominant direction of motion), of which an example is illustrated for surface accelerometer G450 (fig. 3.8). The three-component data are plotted together in three-dimensional space and the ground motion (represented by the position of a virtual particle) is projected onto three perpendicular two-dimensional slices. The results show that in the 49.85 to 50.15 Hz frequency band, the ground motion has a high probability of being on the colored elliptical path and not outside or inside of it, where the probability approaches zero.

Because the polarization was observed to be dominantly in the horizontal plane, the recovered azimuths from the polarization analysis (leftmost panel of fig. 3.8) were projected on geographic maps together with open electrical infrastructure data to identify potential directional sources of the ENF. It was however not possible to identify a regional source of the ENF signal such as medium and high voltage line and

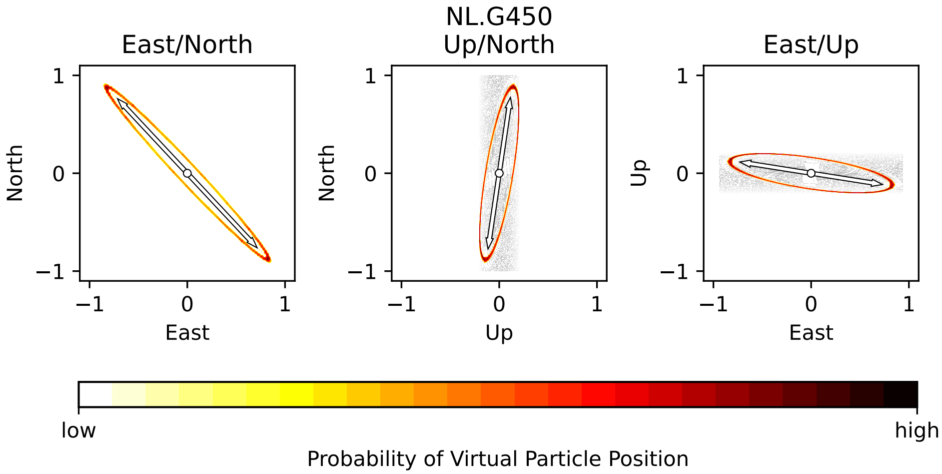

Figure 3.8: *Strongly polarized motion at the ENF (bandpassed* 49.85 *to* 50.15 Hz*) for surface accelerometer G450 (2020-03-01 – 2020-03-02). The three panels show 2-dimensional probabilistic histograms of the particle motion in East/North, Up/North, and East/Up directions respectively. The white arrows with black outline represent the geographic orientation of the largest principal component* ($\mathbf{u}_1$) *that is equal to the direction of dominant particle motion. The leftmost panel can be interpreted as the geographical azimuth of dominant motion.*

transformers. Instead, it was considered that for most instruments, local electronics inside the instrument's housing cabinet may be a more proximal and likely source of the signal. The cabinets that host both the accelerometers and electronics in the G-network is shaped like a rectangular box (ratio 1:3), with the internal setup organised in a similar fashion for all installations. Azimuths of the cabinet in the field (parallel with the elongated side) were estimated from technical drawings. The direction of polarized motion that is expressed by the accelerometer data appears to be consistent with the azimuth of the cabinet (fig. 3.9, left panel), confirming the source of the ENF is in fact local. The right panel of fig. 3.9 shows the misfit between the azimuth of particle motion and the cabinet orientation plotted against the degree of rectilinearity. Stations that express a lower degree of rectilinearity naturally have a larger variability on the direction of particle motion, resulting in a more probable angular misfit. The decreased degree of rectilinearity may be attributed to a diminished source of the ENF or instrument sensitivity issues, which can be considered another instrument health metric.

A clear outlier was identified as station G680 marked in fig. 3.9 that expresses a 87° near perpendicular angular misfit with a very strong rectilinearity (fig. 3.9). It was hypothesised that the instrument was rotated, or that the horizontal components were swapped during instrument installation. A field visit confirmed that surface accelerometer G680 was in fact rotated counter-clockwise by 90° and has been corrected since.

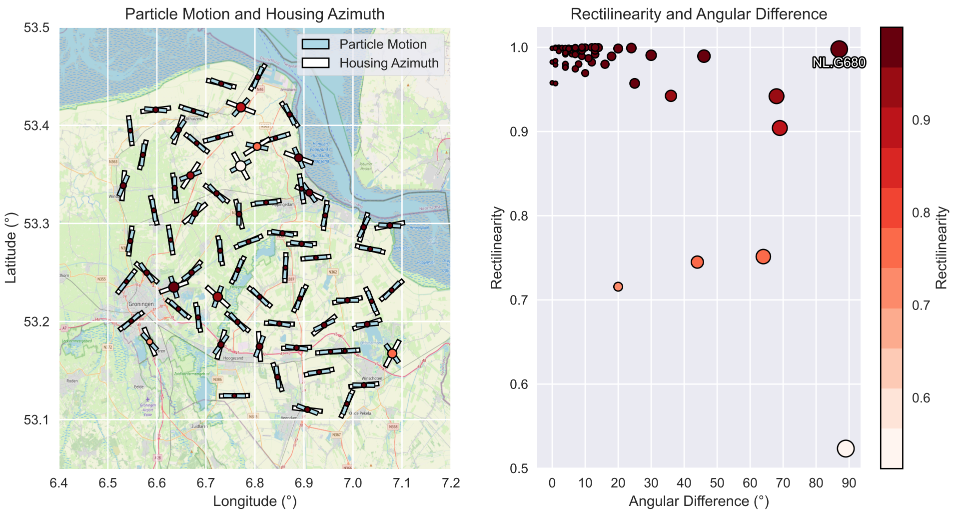

Figure 3.9: ***Left:*** *comparison between azimuths of the principal direction of accelerometer particle motion (lightblue) and the orientation of the installation cabinet (white).* ***Right:*** *Angular misfit between the cabinet orientation and dominant particle motion against the degree of rectilinearity. Map data are provided by OpenStreetMap contributors [2017].*

3.5 Discussion

The applied tracing algorithm (fig. 3.4D) to estimate the ENF from spectrograms using the maximum PSD per time bin is simple yet effective. The intensity of the ENF above ambient noise does not require the use of advanced track tracing algorithms (e.g., *Lampert and O'Keefe* [2010]). For the applications where the ENF needs to be eliminated from the data, subtraction algorithms [*Butler and Russell*, 1993, 2003] may benefit from using reference ENF data too. This is particularly true for extremely (ELF) and very low frequency (VLF) radio data between 300 to 30 000 Hz [*Cohen et al.*, 2010], since the affected bandwidth of the ENF fluctuations grows proportionally with higher overtones. With reference data, the ENF can be specifically targeted and generic bandstop filters can be avoided.

Cross correlations between estimated and reference ENF data provide a reliable, passive technique for the detection of timing anomalies in geophysical data. However, the limitations of the method are clear: the reliability of the timing corrections is contingent on the ability to accurately resolve the ENF signal from the data – which is not always easily achieved. The expected precision and accuracy of the technique illustrated in fig. 3.6 and reaches approximately 1 s for instruments that express a high susceptibility to the ENF. By increasing the sampling resolution of the reference ENF data, time discrepancies on the sub-second level may potentially be discovered. During the analysis of the teleseismic event (fig. 3.7) it was found that there was a consistent delay (6 s) with the reference ENF for the entire NSAN. This delay is not real considering most of the stations are GPS locked and show 0 s delays during other periods. It is expected that this effect may be introduced by poor timing quality of the reference ENF data itself, or potentially by another unknown cause that needs to be investigated further. A similar explanation concerning inaccurate timestamping of the reference ENF data may also explain the skew towards −1 s in fig. 3.6. It should be noted that if the absolute timestamp of the reference data is inaccurate, relative timing differences between instruments using the ENF remain resolvable.

The polarization analysis (figs. 3.8 and 3.9) shows that gross orientation anomalies can be successfully identified. Even if the source of the ENF signal is unknown, when the source remains stable through time (e.g., a non-mobile transformer or the installation cabinet), the rectilinearity of geophysical data at the ENF may thus provide a reasonable tool for the detection of temporal instrumental orientation anomalies. Furthermore, this method may provide a tool to more accurately determine three-dimensional orientations of geophones installed in seismic boreholes that needs to be investigated. Perhaps, even small orientation anomalies may be discovered that are on the order a few degrees.

3.5.1 Source of the ENF in Geophysical Data

The mechanisms through which the ENF signal is passed on to geophysical sensor networks remains enigmatic and appears to vary per instrument type and installation (fig. 3.1). In the following section, the expected sources in the different geophysical instruments are discussed. Because of the alternative suspected coupling mechanism, acoustic instruments are treated in the supplementary information.

G-network Accelerometers and Geophones

From the presented polarization analysis it is evident that the ENF signal is acquired locally in the G-network accelerometers. Despite this, in the operational NSAN, a sudden increase in the amplitude of the ENF has been observed to lead to false event detection in accelerometers deployed near high voltage power lines – suggesting that large-scale electrical infrastructure may under certain circumstances be a significant source of the ENF signal. Seismoacoustic coupling (e.g., *Evers et al.* [2007]) from humming and corona discharge [*Loeb*, 1965] may provide a coupling mechanism up to 200 m away from high voltage power lines [*Schippkus et al.*, 2020]. The susceptibility of the G-network accelerometers to the ENF is strong and highly polarized. It is expected that the signal would be less dominant if it were induced along the wires between the sensor and digitizer where it is not amplified to such dominant amplitudes. Furthermore, accelerometers in the NSAN are connected with a two-wire differential setup, effectively limiting the influence of external magnetic fields on grounds loops specifically, but leaving the sensor itself susceptible to changing magnetic fields. The recorded power at the ENF in accelerometers with different gain settings and sensitivities appears similar across the G-network when the amplitude of the signal is expressed in physical ground motion units (acceleration, velocity, or displacement), suggesting that the ENF signal is not electromagnetic of nature. Alterations in the suspension spring or coils of the accelerometers [*Forbriger*, 2007] have been suggested as a likely source of the signal. The relationship between the cabinet orientation and the polarization azimuth of the accelerometer data indicates that physical vibration of the cabinet itself may be caused by the humming power supply that is mounted on its inside wall.

The geophones inside the seismic boreholes of the G-network share surface electronics with the aforementioned accelerometers. The geophones operate passively and have no direct power source but are connected to the power grid through a digitizer at the surface. The amplitude of the ENF in these data is orders of magnitude smaller compared to the accelerometers and show varying directions of polarization within a single borehole. The polarization is strong, yet orientations vary unpredictably over the 50 m depth levels inside the borehole, and because no decrease with depth inside the boreholes (from 50 to 200 m) could be identified, it is suggested that the ENF signal is potentially established at the surface. For these instruments, it may be that unshielded signal cables connected to the datalogger allow for direct induction of stray magnetic fields from nearby electrical components. A more thorough assessment of the ENF signal in geophones inside the seismic boreholes is recommended.

E-TEST Battery Operated Geophones

Surface geophones from the E-TEST temporary deployment [*Shahar Shani-Kadmiel et al.*, 2020] are fully battery operated and enclosed within a single unit. These instruments are of interest because they have no physical connection to the electrical grid. For these geophones, the ENF signal is only detectable and usable when the instruments are deployed near towns (fig. 3.10), visible overhead power lines, or subsurface electrical infrastructure, as revealed by the presence of e.g., street lights. In the middle of a forest or field, the ENF signal could not be recovered from the data. It is still unknown whether the coupling is purely electromagnetic or through (coupled) waves as a result of the humming and vibration of the nearby electrical components.

Microbarometers

Infrasound sensors express a different sensitivity to the ENF compared to seismometers. Station DBN08 (fig. 3.11) installed at De Bilt expresses the highest sensitivity to the ENF at twice the AC frequency at 100 Hz. Extracting the ENF is more challenging due to the additional anthropogenic noise from a nearby highway reduces the perceptibility of the ENF during the daytime period (05:00 – 21:00), as the amplitude of ENF does not exceed the ambient noise level. The ENF signal appears with a delay of −7 s on infrasound sensor DBN08, suggesting there exists a measurable time lag between the reference and measured ENF. The infrasound sensor uses GPS for time and the recovered delay is not expected to be a timing issue. Hyperion

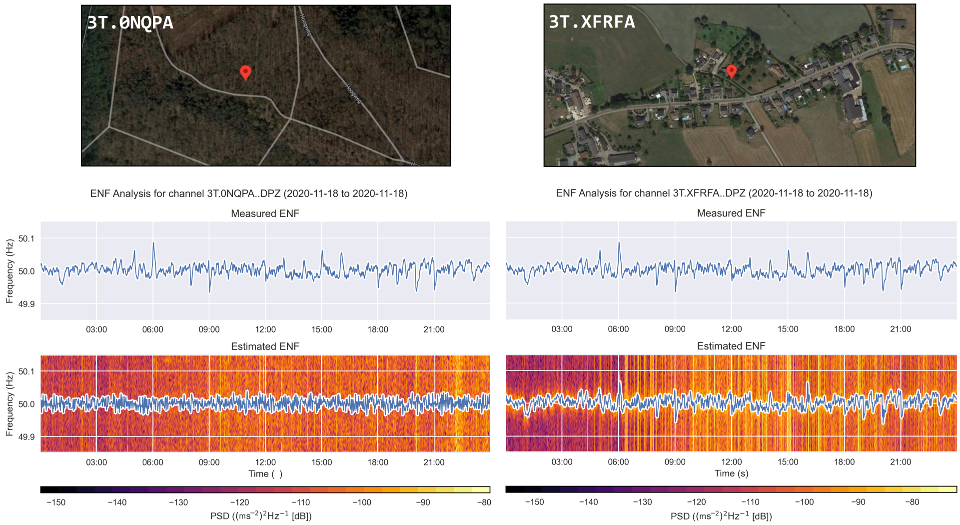

Figure 3.10: *Columns showing two battery operated geophones in the 3T temporary deployment [Shahar Shani-Kadmiel et al., 2020]. The left column shows geophone node 0NQPA remotely deployed in a forest and shows no trace of the ENF in its data. The right column represents data from geophone node XFRFA which is located near a town and electrical infrastructure. The ENF signal is clearly derived from anthropgenic activity in this area.*

microbarometers instruments are specifically designed to limit the influence of external magnetic fields. However, they may still be susceptible to acoustic background noise from vibrating transformers caused by magnetostriction of the transformer's core [*Weiser et al.*, 2000; *Gange*, 2011], in other words, humming. This hypothesis matches the observation of the largest sensitivity at 100 Hz instead of the nominal AC frequency of 50 Hz, where the signal is not clearly detectable. The recovered time delay from the ENF analysis may thus be attributed to the physical travel time from the source (i.e., transformer) to receiver, and should be cautiously interpreted in terms of timing errors. Considering this, the delay may potentially represent an observed travel time of an acoustic wave at the speed of sound ($343\,\mathrm{m\,s^{-1}}$) from a humming transformer at approximately 2 to 2.5 km. If the coupling mechanism is indeed acoustic of nature, the ENF signal may provide a continuous and reliable stable source for acoustic arrays and find applications in interferometry [*Fricke et al.*, 2014].

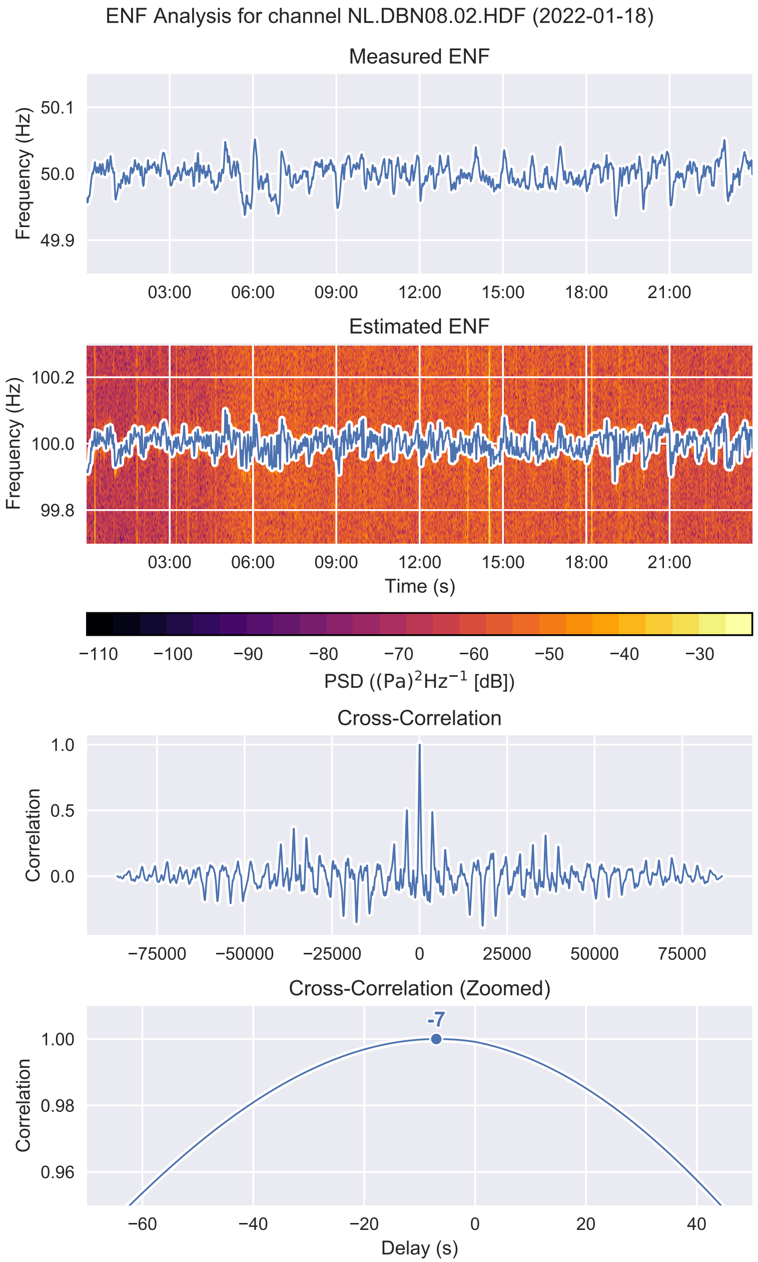

Figure 3.11: *Hyperion infrasound sensor DBN08 (De Bilt, The Netherlands) showing the reference (row 1) and spectrogram-estimated ENF (row 2). Note that the observed signal is at double the nominal AC frequency* ($100\,\mathrm{Hz}$)*. The bottom rows show the resulting cross-correlation between the reference and estimated ENF, providing a measure of time delay* ($-7\,\mathrm{s}$)*.*

■ 3.5.2 Further Applications of ENF Analysis

In the previous sections, the benefits and versatile application of ENF analysis in the passive quality assessment of geophysical data was demonstrated. Because the signal is persistent and omnipresent, some other foreseeable applications and possibilities for future consideration are discussed below.

Seismometers are considered to be linear time-invariant (LTI) systems. This description implies that an input of particular frequency should output a signal with equal frequency, albeit with modified amplitude and phase, as described by the instrument's transfer function. Because the input signal of the ENF is well-defined and predictable, its characteristics should be accurately reflected in the output signal. A number of LOFAR stations in the NSAN network show an anomalous consistent positive shift in the ENF of 0.01 Hz. This feature may represent a deviation from a linear response, or that there exists a minor drift in the clock that may stretch sample spacing, providing the appearance of a higher frequency input signal. The latter hypothesis seems most likely considering the stations are known to use non-commercial dataloggers.

Additionally, the absolute (integrated) amount of power of the observed ENF in digital recordings varies significantly as a function of time. Many features are expressed in this variation, most of which do not yet have identified sources. The most coherent changes happen on timescales of minutes to days and occur simultaneously and proportionally between all stations in the network. Diurnal variation of the strength of the ENF signal appears to be to some degree coherent with measures of the consumer load. An in-depth investigation on these varying amplitude, including a better understanding of coupling mechanisms in geophysical instruments, may provide opportunities for other potential benefits of ENF analysis to be identified, such as the potential detection of sensitivity anomalies. Furthermore, the coherency of the varying ENF signal strength between stations may provide an alternative way to detect relative timing issues that needs to be investigated.

3.6 Conclusion

The application of ENF analysis to the passive quality assessment of geophysical data is a versatile technique that can be leveraged to identify timing issues at the 1 s level. It is also demonstrated that a polarization analysis of accelerometer data at the ENF enabled instrumentation orientation errors to be detected and resolved. ENF analysis may thus be considered for the passive detection of timing errors and sensor orientation anomalies, and in data where the provided timestamp may be tampered with, or generally unreliable, for example due to the lack of GPS connectivity. The mechanism through which the ENF is coupled to geophysical data appears to be instrument and installation specific and needs to be investigated further. Despite this, the proposed methods can potentially be adopted by geophysical monitoring institute, and opens multiple avenues for further research.

3.7 Data and Resources

Reference ENF data were downloaded from the power-grid frequency database [*Gorjão et al.*, 2020] and the TransnetBW GmBH website that is accessible at https://www.transnetbw.com. Seismological waveform data were downloaded from the Netherlands Seismic Acoustic Network [*KNMI*, 1993] and [*Shahar Shani-Kadmiel et al.*, 2020]. The ENF analysis script was written in Python 3.8.2 [*Van Rossum and Drake*, 2009], using SciPy [*Virtanen et al.*, 2020] and NumPy [*Harris et al.*, 2020]. Figures were made with Matplotlib [*Hunter*, 2007], version 3.2.1 [*Caswell et al.*, 2020] and a pre-release version of PyGMT [*Uieda et al.*, 2021] using Generic Mapping Tools (GMT) version 6 [*Wessel et al.*, 2019a,b]. Electrical infrastructure data were downloaded from the Enexis (https://www.enexis.nl/) and TenneT (https://www.tennet.eu) homepages.

3.8 Acknowledgements

We are grateful for our dear colleagues at the Seismology and Acoustics department at KNMI for their fruitful comments and assistance. Two anonymous reviewers are thanked for providing their input on the manuscript and suggestions to improve its clarity. We thank Carrick Talmadge for discussions on the detection of 100 Hz signals on low-frequency sound microphones. Henrietta Rákóczi is thanked for her input on the sections of the manuscript that were considered redundant and could be removed. L. Gorjão is thanked for an interesting discussion on power-grid dynamics and their effort in the creation and maintenance of the power-grid frequency database. This work was initially inspired by a science video created by Tom Scott on the electrical network frequency.

4

Microgravity Change During the 2008 – 2018 Kīlauea Summit Eruption: Nearly a Decade of Subsurface Mass Accumulation

Abstract Results from nine microgravity campaigns from Kīlauea, Hawaiʻi, spanning most of the volcano's 2008 – 2018 summit eruption, indicate persistent mass accumulation at shallow levels. A weighted least squares approach is used to recover microgravity results from a network of benchmarks around Kīlauea's summit, eliminate instrumental drift, and restore suspected data tares. A total mass of 1.9×10^{11} kg was determined from these microgravity campaigns to have accumulated below Kīlauea Caldera during 2009 – 2015 at an estimated depth of 1.3 km below sea level. Only a fraction of this mass is reflected in surface deformation, and this is consistent with previously reported discrepancies between subsurface mass accumulation and observed surface deformation. The discrepancy, amongst other independent evidence from gas emissions, seismicity, and continuous gravimetry, indicate densification of magma in the reservoirs below the volcano summit. This densification may have been driven by degassing through the summit vent. It is hypothesised that during the final years of the summit eruption, magma densification resulted in a buildup of pressure in the reservoirs that may have contributed to the lower East Rift Zone outbreak of 2018. The observed mass accumulation beneath Kīlauea could not have been detected through other techniques and illustrates the importance of microgravity measurements in volcano monitoring.

Koymans et al., (2022). Microgravity Change During the 2008 – 2018 Kīlauea Summit Eruption: Nearly a Decade of Subsurface Mass Accumulation. Journal of Geophysical Research: Solid Earth, e2022JB024739, doi: 10.1029/2022JB024739

4.1 Introduction

Kīlauea, Hawaiʻi (4.1), is one of the most active and accessible shield volcanoes in the world. The volcano provides exceptional opportunities to investigate subsurface magma storage areas and their spatiotemporal evolution. In 2008, a decade-long summit eruption began, characterized by a lava lake that rose and fell according to changes in magma pressure. Effusive activity had also been ongoing since 1983 near the Puʻuʻōʻō vent, 20 km down-rift of the summit on the volcano's East Rift Zone (ERZ). Both eruptions ended suddenly in May 2018, when a dike propagated to the lower ERZ, 40 km down-rift of the summit, erupting over 1 km^3 of material over a period of three months. The outbreak in the lower ERZ left multiple neighborhoods destroyed and was accompanied by the piecemeal collapse of Kīlauea's summit [*Neal et al.*, 2019].

The lava lake that was present during Kīlauea's 2008 – 2018 summit eruption was connected to a shallow magma storage area known as the Halemaʻumaʻu Reservoir (HMMR), centered at 0 to 1 km depth below sea level (b.s.l.). A second, larger magma reservoir exists beneath the south part of the caldera (fig. 4.1) centered 2 to 4 km depth b.s.l. and is referred to as the South Caldera Reservoir (SCR) [*Lundgren et al.*, 2013; *Poland et al.*, 2014; *Jo et al.*, 2015; *Wieser et al.*, 2020; *Bemelmans et al.*, 2021]. Volume estimates of the HMMR range from 0.2 to 5.5 km^3 [*Anderson et al.*, 2015] and 2.5 to 7.2 km^3 [*Anderson et al.*, 2019], and the volume of the SCR is thought to exceed 10 km^3 [*Poland et al.*, 2014]. The level of the lava lake varied synchronously with ground deformation, indicating a strong – possibly *magmastatic* (hydrostatic but with magma) – connection with the HMMR [*Lundgren et al.*, 2013; *Anderson et al.*, 2015; *Orr et al.*, 2015; *Patrick et al.*, 2019a,b; *Poland et al.*, 2021a]. The connection and interplay between the HMMR and SCR remains uncertain [*Poland*, 2014; *Anderson et al.*, 2020; *Wang et al.*, 2021; *Poland et al.*, 2021b].

Summit eruptive activity between 2008 and 2018 was captured by a diverse set of geodetic observations, including campaign and continuous GNSS (Global Navigation Satellite System), borehole tilt, InSAR, and campaign and continuous microgravity [*Poland et al.*, 2021b]. Deformation data are an effective tool for investigating changes in subsurface volume and pressure beneath Kīlauea [*Jo et al.*, 2015; *Bemelmans et al.*, 2021]; however, only gravity measurements can provide constraints on whether surface deformation is accompanied by subsurface changes in mass – information that is critical for understanding the source of geodetic change [*Carbone et al.*, 2017; *Nikkhoo and Rivalta*, 2022]. Mass accumulation may be a precursor to volcanic activity [*Rymer*, 1994], making the technique valuable and widely adopted in volcano monitoring to produce quantitative estimates of source characteristics at volcanoes worldwide [*Carbone and Greco*, 2007; *Miller et al.*, 2017].

4.1.1 Microgravity Observations at Kīlauea

Over the course of the 2008 – 2018 eruption, continuous microgravity monitoring near the summit eruptive vent yielded constraints on the density of the lava lake and dynamics of subsurface magmatism. These constraints were based mostly on short-lived signals that spanned minutes to days [*Carbone and Poland*, 2012; *Carbone et al.*, 2013; *Poland and Carbone*, 2016, 2018; *Poland et al.*, 2021a]. In continuous observations from mechanical spring gravimeters, the contributions of instrumental drift and long-term mass changes are challenging to separate. Furthermore, continuous observations from a single location are insufficient to derive quantitative estimates of the location and magnitude of subsurface mass variations. Microgravity campaigns completed over the course of multiple years overcome these limitations by measuring microgravity at a network of benchmarks relative to a reference benchmark that is outside the area of expected gravity change. These measurements, expressed relative to this reference, can be corrected for instrumental drift by repeating the measurements in a short period over which the drift can be quantified and eliminated [*Van Camp et al.*, 2017]. Changes in the difference between the benchmarks and the reference benchmark can then be observed over longer periods of time. The increased spatial coverage from microgravity campaigns thus allows the source of long-term subsurface mass variations to be resolved. In campaign gravimetry, however, one challenge is aliasing, meaning that short-duration signals during the campaign may be mistaken for changes occurring over longer periods. Together, continuous and campaign microgravity observations provide unique spatio-temporal constraints on subsurface magma dynamics. It is evident that both types of microgravity data are uniquely valuable in volcano monitoring infrastructure.

Microgravity campaigns began at Kīlauea with sporadic measurements in the 1970s [*Jachens and Eaton*, 1980; *Johnson et al.*, 2010] and were completed episodically during the 2008 – 2018 summit eruption using a pair of Scintrex CG-5 gravimeters [*Scintrex Limited*, 2012]. Interpretations based on campaigns completed between 1975 and 2012 suggest significant subsurface mass accumulation below Kīlauea's summit caldera without a commensurate increase in volume, as indicated by a lack of expected surface inflation [*Johnson et al.*, 2010; *Bagnardi et al.*, 2014]. The gravity increase has been interpreted as filling of subsurface voids prior to the onset of the summit eruption in 2008 [*Johnson et al.*, 2010], and the densification of gas-rich magma following the start of summit eruptive activity [*Bagnardi et al.*, 2014].

4.1.2 Microgravity Campaigns and Analysis

Relative spring gravimeters belong to the class of instruments that are sensitive to µGal ($1\,\mu\text{Gal} = 1 \times 10^{-8}\,\text{m/s}^2$) variations in the vertical component of gravitational acceleration g, informally referred to as *microgravity*. In campaign gravimetry, this measure of microgravity is expressed relative to another observed value at a reference benchmark called an *anchor*. Portable gravimeters (e.g., Scintrex CG-5 and CG-6, LaCoste & Romberg Model D and G, ZLS Burris instrument) are practical for field surveys due to their compact size and limited weight. Manufacturer specifications for these instruments suggest a resolution of about 1 µGal and standard deviation of 5 µGal for the Scintrex CG-5 [*Scintrex Limited*, 2012] for repeated measurements. In campaign gravimetry, the term *benchmark* describes a persistent and precise geographical location that is occupied during multiple campaigns. An *occupation* is a single visit to a benchmark during which multiple microgravity measurements are made. During a full campaign that may span days to weeks, multiple *circuits* across a network of benchmarks are completed. Within a circuit, the first measurement is made at the anchor, after which occupations at a number of benchmarks are completed. A high spatial density of benchmarks allows for circuits to be closed, meaning that the anchor is remeasured at the end of each circuit, completing a *loop*. When the same circuit is completed twice, it is referred to as a double loop. The double loop procedure offers insight into measurement repeatability and delivers an important degree of data redundancy. At least one repeated measurement at the network anchor is required to correct for the effect of instrumental drift that is inherent to mechanical spring gravimeters. Data *tares* are sudden and irreversible offsets in gravimeter readings caused by mechanical and/or thermal shocks (e.g., jolting of the instrument by rough transport). This constant offset remains in the data for following occupations. The campaign strategy [*Murray and Tracey*, 2001] can be chosen based on the network topology with a trade-off between time and data redundancy.

Microgravity differences between the benchmarks and the network anchor are calculated for each campaign and compared between campaigns to extract changes in microgravity over time. Many software packages have been developed to process microgravity data, including e.g., GRAVNET [*Hwang et al.*, 2002], MCGravi [*Beilin*, 2006], GTOOLS [*Battaglia et al.*, 2012, 2022], GravProcess [*Cattin et al.*, 2015], PyGrav [*Hector and Hinderer*, 2016], Gsolve [*McCubbine et al.*, 2018], pyGABEUR-ITB [*Wijaya et al.*, 2019], and GSadjust [*Kennedy*, 2021]. The large variety of packages may be a result of lack of standardisation in data collection and analysis, particularly due to different strategies to correct for tidal variations and instrumental drift. In this paper, a custom joint weighted least squares (WLS) inversion is adopted to simultaneously solve for instrumental drift and microgravity differences [*Reilly*, 1970; *Hwang et al.*, 2002]. The approach is extended to correct for microgravity offsets introduced by suspected data tares [*Koymans*, 2022a].

In this manuscript, results from nine microgravity campaigns completed between 2009 – 2017 are presented. The results provide additional constraints on the amount and depth of subsurface mass accumulation beneath the Kīlauea Caldera during its summit eruptive activity.

4.2 Methodology

At Kīlauea, relative microgravity observations were made during nine campaigns [*Flinders et al.*, 2022] by the United States Geological Survey Hawaiian Volano Observatory (USGS – HVO) between 2009 and 2017 (fig. 4.2). Each campaign was carried out using two Scintrex CG-5 instruments (serial numbers 578 & 579) for all benchmark occupations. All circuits across the network (fig. 4.1) were completed as double loops. The benchmark P1 northwest of Kīlauea's summit was used as the network anchor, consistent with previous campaign microgravity studies [*Johnson et al.*, 2010; *Bagnardi et al.*, 2014]. For loops that include benchmarks inside Kīlauea Caldera, benchmark HVO41 was sometimes used as a proxy anchor (fig. 4.1), because logistical constraints made it difficult to return to network anchor P1 between loops. Results from these circuits are therefore expressed relative to the original

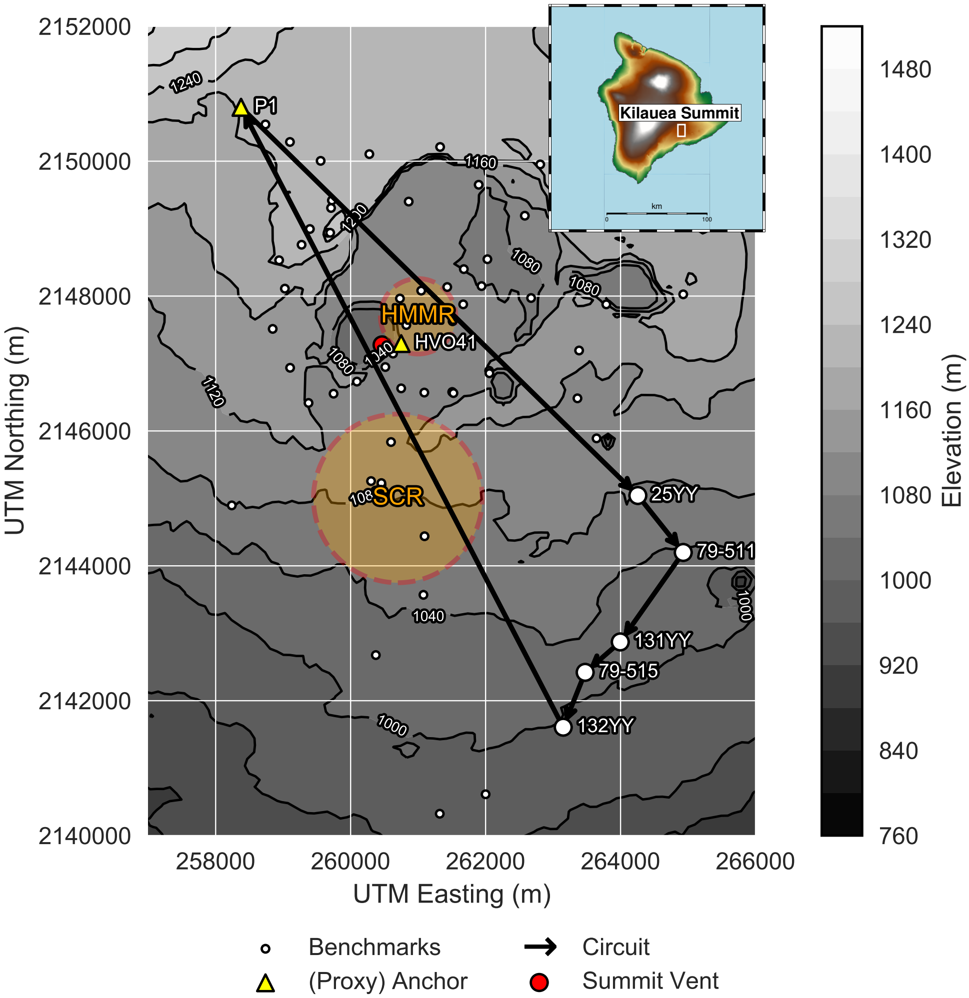

Figure 4.1: *Overview of the summit of Kīlauea, Hawai'i (pre-2018). Each small white circle represents a microgravity benchmark that is made up of a permanently marked site. An example circuit completed on 2010-07-02, visiting a selection of benchmarks (inflated white circles) is illustrated (P1 → 25YY → 79-511 → 131YY → 79-515 → 132YY → P1). After completing a single loop, the circuit is immediately repeated a second time. The approximate projected surface positions of the Halema'uma'u reservoir (HMMR) and South Caldera reservoir (SCR) are shown following the illustration by Poland et al. [2021b].*

anchor through another circuit that captures the difference between network anchor P1 and proxy anchor HVO41 on a different day of the campaign. During every occupation at a benchmark, a minimum of five observations were made, where each observation consists of a 60 s measurement sampled at 6 Hz. The resulting 360

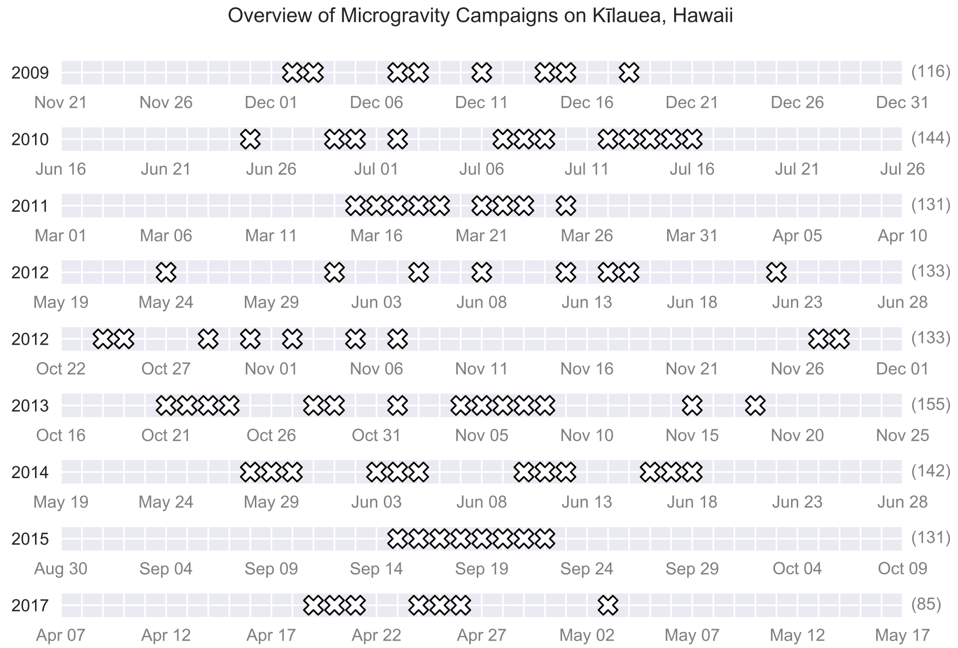

Figure 4.2: *Overview of relative microgravity campaigns on Kīlauea, Hawai'i between 2009 and 2017. The rows span 41 days and each white marker represents a single day on which a circuit across the network was conducted. The number in parenthesis on the right is equal to the number of occupations made during the campaign. A total of nine campaigns were completed with a total of 87 days spent in the field.*

samples are averaged to yield a statistical mean and variance per observation.

The microgravity data from all nine campaigns were uniformly analysed. Initially, clearly erroneous measurements were manually filtered from the data set, including initial observations during each occupation for which the instrument was recovering from transport and converging towards a stable value [*Reudink et al.*, 2014]. The period of recovery can last up to 20 min, but data recovered during the settling may still be valuable considering the trade-off between data quality and time. For campaigns where nearby seismic data (HV.NPT..HHZ) [*USGS Hawaiian Volcano Observatory (HVO)*, 1956] were available, microgravity observations distorted by high amplitude inertial signals (e.g., earthquakes) were identified and excluded from processing. Observations with tilts beyond 20″ from vertical and those that were based on fewer than 60 s of recording were removed. The embedded tidal correction applied by the Scintrex CG-5 software [*Longman*, 1959] was restored to all microgravity observations because of partially erroneous timestamping that propagated to inaccurate tidal corrections in some of the campaigns. The effect of the solid Earth tide was again removed using a branch of Pygtide [*Rau*, 2018], a Python wrapper for ETERNA 3.4 [*Wenzel*, 1996]. The tidal components were estimated using a global theoretical model following *Dehant et al.* [1999] as recommended by *Van Camp and Vauterin* [2005] in the TSOFT manual. The effect of ocean loading was removed using parameters obtained from the free ocean loading provider [*Bos and Scherneck*, 2014] using the TPXO9-atlas [*Egbert and Erofeeva*, 2002]. The tidal constituents returned from the provider were evaluated using the IERS standard program HARDISP [*Agnew*, 2010].

During a single circuit spanning a period of hours, instrumental drift of the CG-5 can be characterized by a monotonic linear function. The instruments implement an automated drift correction, but a residual drift always remains to be considered. A weighted least squares inversion was used to find a solution for the gravity differences with the anchor and linear instrumental drift parameters [*Hwang et al.*, 2002], including any offsets introduced by suspected data tares [*Koymans*, 2022a]. An example solution for a single circuit is illustrated in fig. 4.3. A detailed derivation of the microgravity analysis and uncertainty estimation is provided in appendix A.

4.2.1 Deformation Correction

InSAR data were acquired by the X-band TerraSAR-X and COSMO-SkyMed satellite systems (table 4.1). Data were processed using the GAMMA software [*Werner et al.*, 2000], with the topographic correction made using a 5 m digital elevation model [*Poland*, 2014]. Satellite line-of-sight displacement vectors $\mathbf{U}^*$ were calculated from unwrapped interferograms $\mathbf{U}$ following the same approach as *Bagnardi et al.* [2014]. The satellites are, by approximation, characterized by north-to-south orbital trajectories, with a heading that is off by a maximum of a few degrees. By this approximation, the line-of-sight displacement vectors from the ascending and descending traces can be decomposed to a horizontal (east to west) and vertical component [*Yun et al.*, 2006]:

$$\begin{bmatrix} \mathbf{U}^*{}_{\text{descending}} \\ \mathbf{U}^*{}_{\text{ascending}} \end{bmatrix} = \begin{bmatrix} \sin(\lambda_1) & \cos(\lambda_1) \\ -\sin(\lambda_2) & \cos(\lambda_2) \end{bmatrix} \begin{bmatrix} \mathbf{U}_{\text{z}} \\ \mathbf{U}_{\text{h}} \end{bmatrix} \tag{4.2.1}$$

Where $\mathbf{U}^*$ represents the vector of line-of-sight displacement for the respective trajectory, and λ_1 (descending) and λ_2 (ascending) the corresponding satellite incidence angles measured from horizontal. Given the line-of-sight displacements for the ascending and descending trajectory, Equation 1 can be used to find solutions for the vertical ($\mathbf{U_z}$) and horizontal (east to west) ($\mathbf{U_h}$) deformation. Despite the occasional mismatch in temporal coverage, ascending and descending interferograms are acquired within a few days of one another. Furthermore, the interferograms span months to years and were chosen without overlap with respect to known major deformation events, like the 2011 Kamoamoa fissure eruption [*Lundgren et al.*, 2013] and 2015 summit intrusion [*Jo et al.*, 2015; *Bemelmans et al.*, 2021]. Vertical deformation data were smoothed using a median filter over an area of roughly 100 m.

Vertical deformation estimates from InSAR data were constrained using continuous GNSS data by minimising the squared residual between the InSAR and

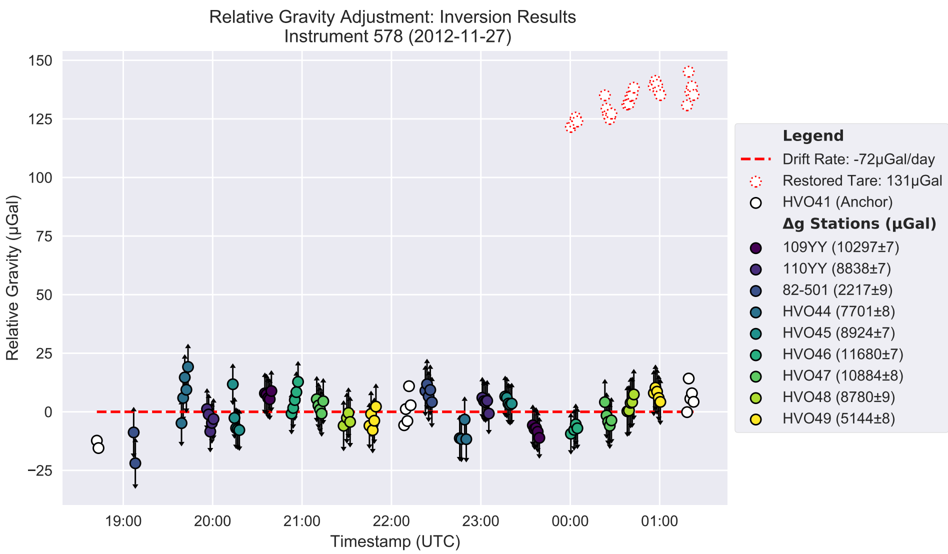

Figure 4.3: *Observations from a single campaign day (2012-11-27) using Scintrex CG-5 578 on Kīlauea, Hawai'i. The vertical error bars illustrate the* 95 % *confidence interval for the recovered gravity differences with the anchor. Proxy anchor HVO41 has no confidence interval because the data are kept fixed during the inversion. Solutions for the gravity differences with the anchor (subtracted from observations) and linear instrumental drift (corrected) are shown. The magnitude of the recovered microgravity differences and uncertainties are given in the legend. The residual daily linear drift rate is estimated at* −72 µGal/d. *An additional degree of freedom that represents a data tare was added to the inversion for the group of data marked in red. This tare was restored during the inversion with an offset of* 131 µGal.

GNSS data at the GNSS station locations. Estimates of the vertical position of twelve GNSS stations over the periods that match the InSAR acquisition dates were manually interpolated from the plotted time-series data (refer to the online version of this manuscript). The amount of vertical deformation at each benchmark was estimated for the InSAR acquisition dates that closely match the microgravity campaigns. The recovered change in height for each benchmark relative to network anchor P1 was converted to gravity assuming a theoretical free-air gradient of −308 µGal/m [*Fowler et al.*, 1990]. Measured local gradients of −327.3 µGal/m [*Johnson*, 1992] and −330.25 µGal/m [*Kauahikaua and Miklius*, 2003] are close to this theoretical estimate and should not significantly influence the results [*Johnson et al.*, 2010; *Bagnardi et al.*, 2014] when deformation remains limited to tens of cm.

Table 4.1: *Satellites and temporal coverages used for InSAR vertical deformation estimates.*

		Descending Trace			**Ascending Trace**		
Epoch	X Band Satellite	Start Date	End Date	λ_1 (°)	Start Date	End Date	λ_2 (°)
2009 - 2010	TerraSAR-X	2009-12-04	2010-06-09	31.2	2009-12-05	2010-06-09	33.2
2010 - 2011	TerraSAR-X	2010-06-09	2011-04-02	31.2	2010-06-10	2011-04-03	33.2
2011 - 2012	COSMO-Skymed	2011-03-10	2012-05-15	41.5	2011-03-11	2012-05-28	38.8
2012 - 2012	COSMO-Skymed	2012-05-15	2012-10-30	41.5	2012-05-20	2012-10-23	38.8
2012 - 2013	COSMO-Skymed	2012-10-30	2013-12-04	41.5	2012-10-23	2013-11-27	38.8
2013 - 2014	COSMO-Skymed	2013-11-02	2014-05-29	41.5	2013-10-26	2014-06-20	38.8
2014 - 2015	COSMO-Skymed	2014-05-29	2015-09-06	41.5	2014-06-20	2015-09-02	38.8
2015 - 2017	COSMO-Skymed	2015-09-06	2017-04-23	41.5	2015-09-11	2017-04-16	38.8

4.2.2 Point Source Inversion

Observed microgravity changes that were corrected for vertical deformation can be inverted to recover best-fit point source parameters. These sources are described by a change in subsurface mass and associated position. This geometry has proven to be a suitable approximation for microgravity change sources at Kīlauea's summit in previous studies [*Johnson et al.*, 2010; *Bagnardi et al.*, 2014; *Poland et al.*, 2019] and is often used to model the HMMR [*Lundgren et al.*, 2013; *Poland et al.*, 2014; *Anderson et al.*, 2015; *Bemelmans et al.*, 2021].

For a point source, the vertical change in gravity (δg_i) at the i^{th} benchmark in the Kīlauea network can be approximated by:

$$\delta g_i = \mathrm{G}\delta m \frac{(\delta z - z_i)}{r^3} - \delta g_{\mathrm{P1}} \tag{4.2.2}$$

Where $\mathrm{G} = 6.674 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}$ is the universal gravitational constant, δm the change in mass, and $r = \sqrt{(\delta x - x_i)^2 + (\delta y - y_i)^2 + (\delta z - z_i)^2}$ equals the Euclidean distance between the benchmark position (x_i, y_i, z_i) and the inferred point source location (δx, δy, δz). The term δg_{P1} accounts for the predicted effect of the point source felt at the anchor location and can be found by recursive application of eq. 2. The squared residual between the modeled and observed changes in gravity difference is minimised with respect to the four variable point source parameters (δm, δx, δy, δz) using the Powell method [*Powell*, 1964] implemented in SciPy [*Virtanen et al.*, 2020]. The point sources and gravity residuals were calculated from the average gravity changes of the CG-5 instruments in order to maximise the gravity signal to noise ratio. Three benchmarks near the summit eruptive vent (HOVL-G, HVO41, and 205YY) that are strongly influenced by variations in the lava lake level were excluded from the inversion.

An estimate for confidence intervals on the source inversion results is made through a parametric bootstrap (n = 5000) by sampling the standard deviation of the microgravity changes. From this ensemble of point source solutions, the bottom and top 2.5 % parameter estimates are discarded, resulting in discrete 95 % confidence interval for the point source parameters (δm, δx, δy, δz), including the median of the ensemble of bootstraps.

4.3 Results

The results from the full vertical deformation and microgravity analysis are summarised in fig. 4.4.

■ 4.3.1 Vertical Deformation

The vertical displacement maps in the first row of fig. 4.4 illustrate vertical deformation estimates from the InSAR data that were constrained by the continuous GNSS stations. The vertical displacement from the InSAR data at each microgravity benchmark is shown for comparison. Vertical deformation captured by GNSS and InSAR is consistent within 1 to 2 cm ($\approx$ 3 to 6 µGal), well within the typical 20 µGal uncertainty of microgravity change derived from campaign measurements [*Poland and de Zeeuw-van Dalfsen*, 2019].

Between 2009 and 2010 (fig. 4.4 – row 1, panel 1), a maximum of 3 cm of subsidence occurred south of Kīlauea Caldera, with a negligible amount of deformation within the caldera itself. The subsidence trend continued into 2011, with the 2010 – 2011 period seeing up to 5 cm of subsidence in the center of the caldera (fig. 4.4 – row 1, panel 2). From March 2011 to June 2012, deformation in the caldera was characterized by uplift of up to 15 cm focused on the center of the caldera, while the south part of the caldera remained relatively stable (fig. 4.4 – row 1, panel 3). Over the course of mid to late 2012, no substantial deformation was identified at Kīlauea's summit (fig. 4.4 – row 1, panel 4). Between 2012 and 2014, slow summit uplift began in the caldera center which migrated towards the south (fig. 4.4 – row 1, panels 5 & 6). After 2014, uplift occurred across the summit region but was mainly centered south of the caldera, with uplift rates exceeding 10 cm/yr (fig. 4.4 – row 1, panels 7 & 8) above the expected position of the SCR. The progression of deformation is consistent with GNSS time series that illustrate an overall transition from subsidence to uplift over the course of the 2008 – 2018 summit eruption, interrupted by occasional transient displacements caused by summit and ERZ intrusions and eruptions [*Poland et al.*, 2021b].

■ 4.3.2 Microgravity

The panels presented in fig. 4.4 rows 2 and 3 illustrate microgravity results corrected for vertical deformation at all benchmarks for instruments 578 and 579 respectively.

Between December 2009 and June 2010 (fig. 4.4 – rows 2 & 3, panel 1), an increase in gravity on the order of tens of µGals in the center of the caldera is apparent in data from both gravimeters. There is no coherent pattern from either gravimeter in the subsequent epoch spanning June 2010 to March 2011 (fig. 4.4 – rows 2 & 3, panel 2). A major decrease of approximately 70 µGal during this epoch occurs at a benchmark on the western flank of Kīlauea. This feature is expressed in both instruments but is not representative of the wider area, indicating a local effect or measurement artifact. An increase of similar magnitude happens at a single benchmark to the southeast. From March 2011 to June 2012 (fig. 4.4 – rows 2 & 3, panel 3), an increase in gravity (50 to 200 µGal) appears near the summit eruptive

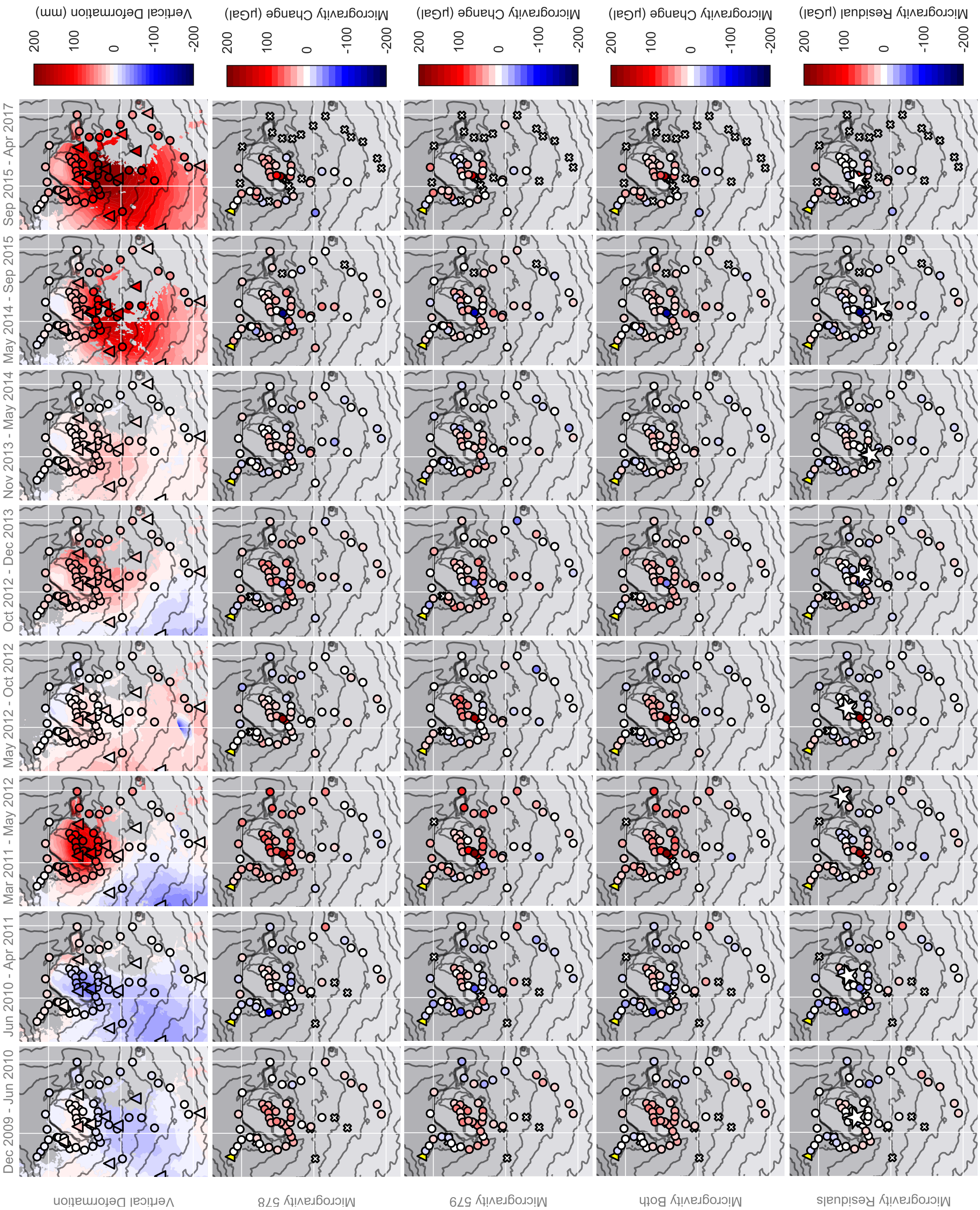
Dec 2009 - Jun 2010
Jun 2010 - Apr 2011
Mar 2011 - May 2012
May 2012 - Oct 2012
Oct 2012 - Dec 2013
Nov 2013 - May 2014
May 2014 - Sep 2015
Sep 2015 - Apr 2017
Vertical Deformation
Microgravity 578
Microgravity 579
Microgravity Both
Microgravity Residuals
Vertical Deformation (mm)
Microgravity Change (μGal)
Microgravity Change (μGal)
Microgravity Change (μGal)
Microgravity Residual (μGal)
200
100
0
-100
-200

Figure 4.4: *Caption for figure on previous page – Columns represent the period between two consecutive microgravity campaigns on Kīlauea, Hawaiʻi. Row 1) vertical deformation (triangles = Global Navigation Satellite System receivers, circles = benchmarks), Rows 2) and 3) gravity changes corrected for vertical deformation for instruments 578 and 579 respectively (triangle = anchor P1, circles = benchmarks, cross = missing). Row 4) average gravity changes for instruments 578 and 579. Row 5) inverted point source solutions and microgravity residuals after correcting for the source (star = source location). The spatial coverage of the panel frames is identical to fig. 4.1 and omitted to save space.*

vent and the general region near the center of Kīlauea Caldera, stretching towards Kīlauea Iki crater and to the southeast. The increase in microgravity is largest at the benchmark in closest proximity to the lava lake (HOVL-G), and similarly elevated for two nearby benchmarks (205YY, and HV041). In data spanning mid to late 2012 (fig. 4.4 – rows 2 & 3, panel 4), the gravity increase near the lava lake persists, with almost all variation (100 to 200 µGal) happening in the vicinity of the eruptive vent. No further significant changes in gravity can be identified during this period for instrument 578, but instrument 579 is characterized by a large increase (40 to 60 µGal) inside Kīlauea Caldera. Between November 2012 and 2013 (fig. 4.4 – rows 2 & 3, panel 5), a gravity increase occurred in Kīlauea Caldera, showing a pattern that is similar in spatial extent and magnitude as that from December 2009 to June 2010. Very little change in gravity can be observed during 2013 – 2014 (fig. 4.4 – rows 2 & 3, panel 6). Between June 2014 and September 2015 (fig. 4.4 – rows 2 & 3, panel 7), the most noteworthy gravity change is a 150 µGal decrease near the summit eruptive vent. The results from 2015 to 2017 (fig. 4.4 – rows 2 & 3, panel 8) are limited in the number of available benchmarks and appear inconclusive, but data from instrument 578 indicate an increase in gravity nearest to the summit vent, with little to no change elsewhere. Row 4 of fig. 4.4 illustrates the average of both instruments and was used as input data for the point source inversions.

Results from the entire period spanning December 2009 to April 2017 are mainly characterized by a persistent increase in gravity that radiates outward from the center of Kīlauea Caldera. The gravity variations with the largest amplitude are observed at the benchmarks near the summit eruptive vent (HOVL-G, HVO41, and 205YY) and closely follow the level of the rising and falling lava lake (fig. 4.8).

Differences Between Gravimeters CG-5 578 and 579

The comparison of microgravity results for the two instruments in fig. 4.4 – rows 2 & 3, illustrates that there is variation in what two different gravimeters record, even though the instruments have been subjected to the same conditions and modes of transport. However, when considering the confidence limits on the mean gravity difference (15 to 20 µGal) that include operator and environmental noise, the instruments generally occupy the same range. The gravity residuals from the linear drift model for each circuit are presented in fig. 4.5. The residuals approximate a normal distribution, indicating that the applied drift model is appropriate. Uncertainties on the mean gravity differences are on the order of a few to a few tens of µGal, depending on the instrument, campaign, and circuit. One implication of the residuals is that gravity changes derived from campaign measurements at Kīlauea cannot be resolved with confidence below approximately 20 µGal. Based on merit of its lower residuals, instrument 578 outperformed instrument 579 (fig. 4.5). Instrument 579 was in fact found to be unreliable in campaigns at Yellowstone during 2017 [*Poland and de Zeeuw-van Dalfsen*, 2019]. These campaigns exposed instrumental problems that may also have been present earlier and could explain the discrepancy between the two instruments observed in the 2015 and 2017 measurements at Kīlauea. This deterioration is not clearly expressed in the gravity residuals in fig. 4.5; however, such residuals would not capture constant offsets from e.g., calibration errors that affect the mean microgravity results. Furthermore, single loops were more commonplace in the 2017 campaign, effectively producing lower residuals but definitely not more reliable data. The instruments were calibrated against absolute measurements on a line at Mount Hamilton, California. The instrument calibration factors did not change between the 2009 – 2017 campaigns, but it was apparent from calibration line measurements that instrument 579 needed servicing in late 2017 [*Battaglia et al.*, 2018].

4.3.3 Recovered Point Source Solutions

Solutions and bootstrapped parameter estimates are summarised in table 4.2 and fig. 4.4 – row 5 and fig. 4.6. In the following paragraphs the results are discussed by epoch:

- **December 2009 to June 2010** (fig. 4.4 – row 5, panel 1) shows a shallow source at 500 m depth b.s.l. towards the northeast of Halemaʻumaʻu crater with a mass change on the order of 2.0×10^{10} kg. The parameter estimates in fig. 4.6 follow Gaussian distributions with tight confidence bounds on the median of the parameter estimates. The location of the point mass is consistent with the observed radial pattern of gravity changes presented in fig. 4.4.

- **June 2010 to March 2011** (fig. 4.4 – row 5, panel 2) resolves to a very shallow depth, with an order of magnitude less mass addition compared to the previous epoch. The distribution of bootstrapped parameters appears skewed and may be attributed to the absence of large and coherent gravity changes during this period.

- **March 2011 to June 2012** (fig. 4.4 – row 5, panel 3) initially resolved to a greater depth and mass addition than preceding epochs. Its resolved location is towards the northeast of Kīlauea Caldera at a location where no large mass change is expected (illustrated in fig. 4.4). Therefore the horizontal position of the point source was kept fixed at the approximated surface location of the center of the HMMR (261000, 2147500) and the change in mass for this source becomes 1.1×10^{11} to 1.6×10^{11} kg at 2.4 to 2.9 km depth b.s.l. The fixed point source is not illustrated in fig. 4.4 but its resolved parameters are provided in table 4.2.

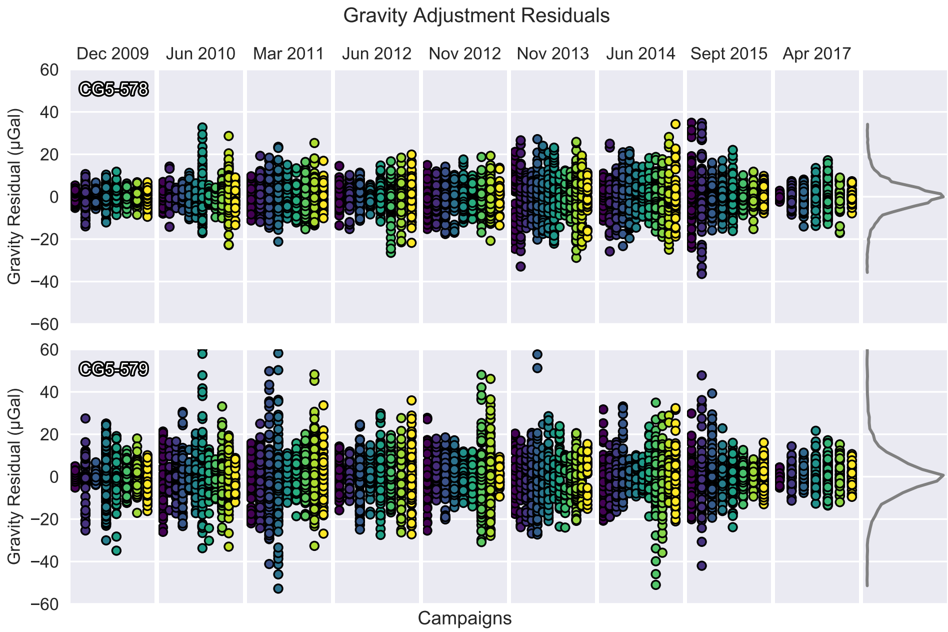

Figure 4.5: *Microgravity residuals from the estimated linear drift model for instruments 578 (top) and 579 (bottom) for all campaigns and circuits on Kīlauea, Hawai'i. Each consecutive circuit in a campaign is given a distinctive color and a new entry on the x-axis. In the ideal case where the model fits the observations the residuals of each circuit should be normally distributed. Deviations from this behaviour may indicate problems with the applied linear drift model, instruments, or the measurements themselves. The right-most panels show histograms of the residuals, illustrating a tighter distribution on instrument 578 compared to 579.*

- **June 2012 to November 2012** (fig. 4.4 – row 5, panel 4) shows a modeled source located towards the center of Kīlauea Caldera. This result is mostly influenced by the data from instrument 579. The ensemble of point sources resolve with Gaussian bootstrapped confidence intervals at shallow depth with little mass change.

- **November 2012 to November 2013** (fig. 4.4 – row 5, panel 5) resolves below Kīlauea Caldera, with a mass change of approximately 7.8×10^{10} kg at 2.5 km depth b.s.l. The gravity increase for this period appears to decrease with distance from the source, and the inversion provides a robust point source solution.

- **November 2013 to June 2014** (fig. 4.4 – row 5, panel 6) shows multiple peaks in the parameter distributions. The optimization thus recovers multiple stable point sources (bi-modality in fig. 4.6), particularly with a depth and mass trade-off. Surface microgravity observations cannot distinguish between an increase in depth or decrease in mass. This may result in multiple distributions of point sources representing greater depth and mass versus shallower depth and smaller mass.

- **June 2014 to September 2015** (fig. 4.4 – row 5, panel 7) indicates a point source at greater depth and larger increase in mass than preceding epochs.

- **September 2015 to April 2017** (fig. 4.4 – row 5, panel 8) expressed bimodal behavior in terms of mass and depth and cannot be reliably resolved. The poor quality of the 2017 campaign and lack of coherent gravity changes for this period make the result unsurprisingly inconclusive.

All modeled point sources indicate mass increase beneath Kīlauea Caldera. Because of the low signal-to-noise ratio of the microgravity observations, results from changes between adjacent campaigns are often inconclusive. To provide a more robust estimate, the gravity changes are integrated over the period spanning December 2009 to September 2015 (omitting the poor 2017 campaign). This point source solution is represented by a mass increase of 1.6×10^{11} to 2.4×10^{11} kg at a depth of 1.0 to 1.7 km b.s.l. beneath the center of the caldera, with Gaussian and narrow confidence intervals (fig. 4.7). The recovered horizontal location is consistent with that of the shallow HMMR, but at slightly greater depth, indicating that perhaps the base of the reservoir is a dominant region of mass accumulation beneath Kīlauea Caldera – a result similar to previous studies [*Johnson et al.*, 2010; *Bagnardi et al.*, 2014; *Poland et al.*, 2019]. The additional results presented here demonstrate that mass accumulation proceeded even during the later stages of the 2008 – 2018 eruption.

4.4 Discussion

The analysis of multiple microgravity campaigns at the summit of Kīlauea spanning 2009 – 2017 provides a foundation for understanding the dynamics of subsurface magmatism, as well as strategies for optimizing the quality and utility of campaign gravimetry. One of the current challenges in terrestrial microgravity exists in increasing the spatio-temporal resolution of data. It is evident that campaign gravimetry is mainly limited by its temporal resolution and high uncertainties caused by often unquantified external effects. This limitation is not inherent to the technique but mainly imposed by high instrumental cost and the time and personnel needs of microgravity campaigns. Projects that aim to surmount these challenges are currently being undertaken [*Carbone et al.*, 2020] by utilizing arrays of low-cost MEMS gravimeters [*Middlemiss et al.*, 2016]. Recommendations from this manuscript may be applied in the future to collect high-quality data targeting magmatic processes that might otherwise remain ambiguous.

Table 4.2: *Median of the point source inversion parameters for all periods between microgravity campaigns at Kīlauea, Hawai'i. The discrete* 95 % *confidence intervals are illustrated in fig. 4.6. The entry marked by an asterisk (*) had its surface location kept fixed. The given depths are expressed relative to sea level (b.s.l. = below sea level). The surface elevation of the Kīlauea Caldera floor is approximately* 1100 m.

	Inverted Source Parameters			
Campaign Period	Mass (kg)	UTM Easting (m)	UTM Northing (m)	Depth b.s.l. (m)
Dec 2009 - Jun 2010	2.0×10^{10}	261000	2147400	0.5×10^{3}
Jun 2010 - Mar 2011	1.3×10^{9}	261500	2148000	-0.6×10^{3}
Mar 2011 - Jun 2012	2.2×10^{12}	264500	2148500	9.1×10^{3}
Mar 2011 - Jun 2012*	1.3×10^{11}	261000	2147500	2.6×10^{3}
Jun 2012 - Nov 2012	5.6×10^{9}	261200	2148000	-0.2×10^{3}
Nov 2012 - Nov 2013	7.8×10^{10}	261200	2147000	2.5×10^{3}
Nov 2013 - Jun 2014	1.2×10^{10}	260100	2136400	1.0×10^{3}
Jun 2014 - Sept 2015	9.5×10^{10}	260900	2145800	3.5×10^{3}
Sept 2015 - Apr 2017	1.2×10^{10}	260400	2147300	-0.2×10^{3}
Overall Period				
Dec 2009 - Sept 2015	1.9×10^{11}	261200	2147400	1.3×10^{3}

4.4.1 Campaign Strategy and Network Adjustment Method

This work demonstrates the effectiveness of a WLS inversion [*Hwang et al.*, 2002] to recover relative gravity differences from double-looped circuits. By utilizing this approach, data tares can be automatically restored when the group of data affected by the tare is known [*Koymans*, 2022a]. Furthermore, instead of using the mean of an occupation, individual measurements are used in the WLS inversion. As a result, the observed drift during a single occupation (which can be up to 1 µGal/min) contributes information to the solution. Finally, an inaccuracy in the solid Earth tide or ocean loading models may not entirely prevent residual higher-order harmonic signals from being present in the measurements. In a circuit that spans up to 12 h, a second- or third-order polynomial may help eliminate any residual (harmonic) components. The linear drift correction in the WLS can be trivially extended to correct for higher-order trends [*Koymans*, 2022a]. Higher-order trends could be identified in a preliminary attempt with a linear drift and subsequently assessing the residuals from this linear model. However, with an insufficient number of occupations a high-order trend may tend to over-fit the data and be detrimental to the results. For this reason, all circuits presented in this manuscript were fitted with a linear drift model.

Single vs. Double Closed Loops

A single occupation (composed of multiple observations) from a benchmark, as during a single-loop circuit, when presented with a degree of freedom, will always align itself with the imposed drift model. When at least two occupations of a benchmark (taken with hours in between) align with the drift model, confidence in the

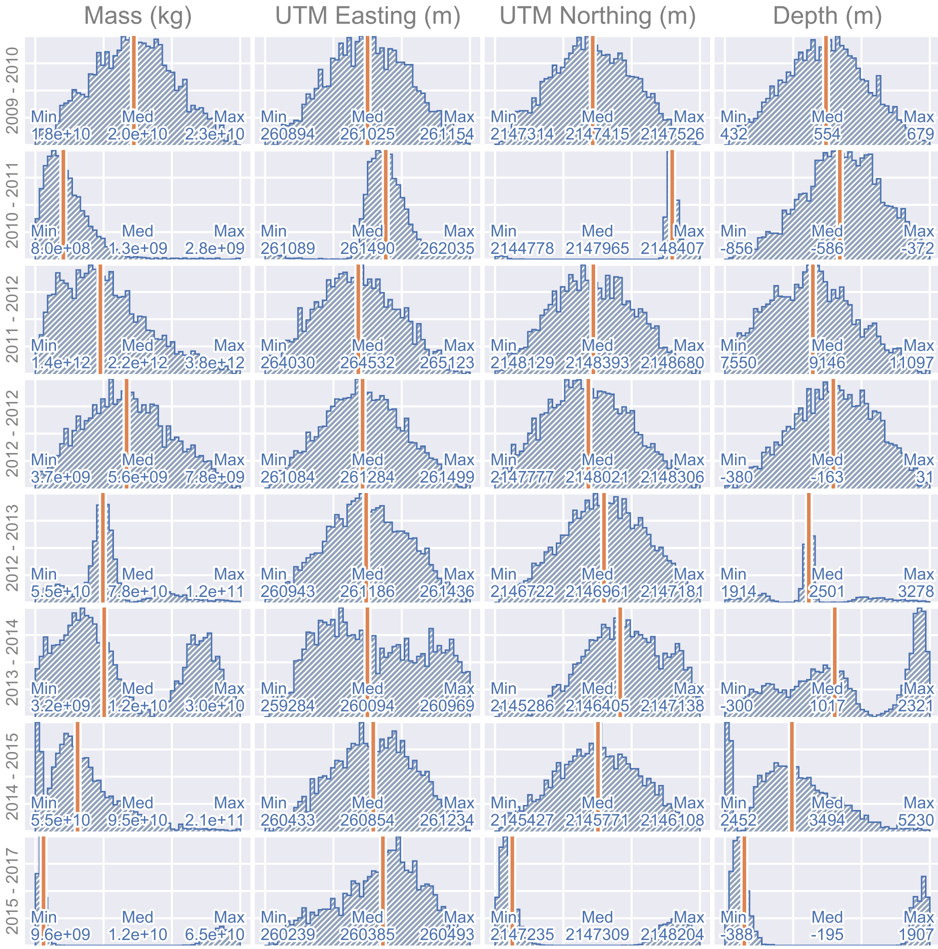

Figure 4.6: *Histograms showing the parameter distributions of 5000 bootstrapped inverted point sources for microgravity campaigns on Kīlauea, Hawai'i for the indicated time periods in table 4.2. The lower 2.5 % and upper 97.5 % discrete percentiles are given as confidence bounds on the median value and define the boundary of the histograms. The median (med) value is represented by the vertical orange bar.*

result increases significantly. This means that multiple occupations of a benchmark provide important insight into measurement repeatability. A single loop is always insufficient because a tare in the data may be misinterpreted as instrumental drift. The presence of tares appears relatively common, with a total of 14 suspected tares observed (offsets between 20 to 130 µGal) over all the Kīlauea campaigns for both instruments combined. The effect of occasionally completed single loops becomes most apparent in the campaigns from 2017, which have low coherence between in-

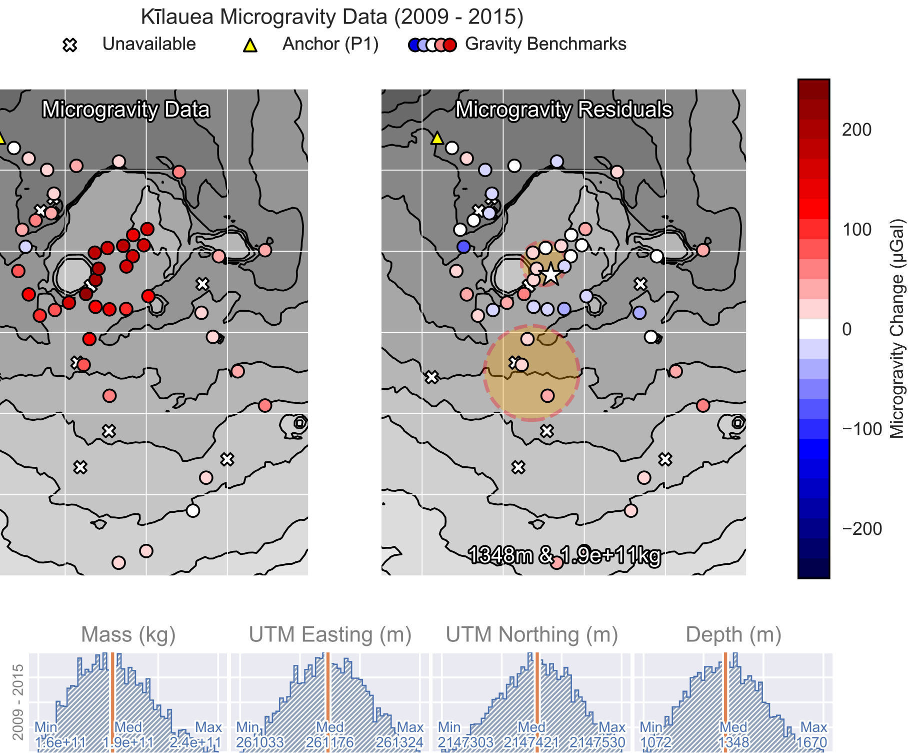

Figure 4.7: *Maps showing the observed gravity changes (left) and inferred point source location and remaining gravity residuals (right) between December 2009 and September 2015 on Kīlauea, Hawai'i. The yellow triangle represents network anchor P1, and benchmarks for which data is missing are marked by a white cross. The location of the inverted point source is illustrated by a star, with an annotation for the associated mass & depth at the bottom. The bootstrapped parameter estimates are shown in the bottom row of panels. The spatial coverage and surface reservoir projections are equivalent to fig. 4.1.*

struments and relatively poor data quality. Due to the inherently low repeatability of microgravity measurements, a certain degree of data redundancy is always recommended. Double loops are clearly favorable in environments that are characterized by significant ambient noise and where rough transport of the instruments cannot be avoided. A key challenge is finding the right balance in the trade-off between effort and data quality, but a minimum of two occupations at each benchmark is needed for reliable results.

The advantage of repeating occupations with multiple instruments is evident from the results and can also assist with the detection of data tares. The presented discrepancies between instruments 578 and 579 emphasize that is it especially valuable when the instruments can be cross-calibrated against an absolute reference before being used in the field [*Miller et al.*, 2017].

Pitfalls of (Proxy) Anchors

Besides instrumental calibration errors, a systemic bias is introduced to the results when microgravity changes occur at the network anchor. An absolute measurement of microgravity may be used to rule out such changes [*Van Camp et al.*, 2017], but was not available for the presented Kīlauea microgravity campaigns. Because the network anchor P1 is located at a site away from volcanic activity, no significant microgravity change or bias in the results is expected, nor observed in fig. 4.4 that would appear as a consistent increase or decrease in all benchmarks.

An important recommendation based on the microgravity results from Kīlauea is to consistently measure all benchmarks in a direct circuit with the network anchor. Each measurement in a circuit has an associated uncertainty and when expressed through a proxy anchor, as was done for some occupations at benchmarks inside Kīlauea Caldera, the uncertainties of both measurements are compounded. More importantly, transient gravity changes caused by variations in the lava lake level affect proxy anchor HVO41. Because the lava level can vary over time scales of hours to days [*Patrick et al.*, 2019b], it is possible that the lava lake level will induce significant differences in gravity at HVO41 on the day that it is used as a proxy anchor for caldera-floor benchmarks compared to the day that it is tied to network anchor P1. In this situation, a bias is introduced when benchmarks originally measured relative to proxy anchor HVO41 are expressed relative to network anchor P1. This complication may explain some of the microgravity-change results between 2010 and 2011, where the observed gravity increase in Kīlauea Caldera (fig. 4.4 – rows 2 – 4, panel 2) may be attributed to the sudden rise in the lava lake level in 2011 over two days. Similarly, a single poor measurement between network anchor P1 and proxy anchor HVO41 may also explain the anomalous increase in Kīlauea Caldera for instrument 579 between June and November 2012 (fig. 4.4 – row 3, panel 4). These artefacts of the campaign strategy may explain why measurements on the caldera floor appear consistently higher during this period, and those data should be interpreted with caution in terms of the modeled source mechanisms. It naturally follows that an overestimation in one campaign would result in an underestimation in the next. Proxy anchors that are subject to high noise and transient microgravity effects should be avoided whenever possible.

4.4.2 Uncertainties in Campaign Gravimetry

Relative microgravity measurements are subject to many sources of uncertainty [*Van Camp et al.*, 2017; *Giniaux et al.*, 2017; *Poland and de Zeeuw-van Dalfsen*, 2019] and are notoriously difficult to interpret. Microgravity campaigns should ideally be completed at regular intervals (e.g., monthly or yearly) and not just in response to disruptive and transient events. All sources of microgravity change between two campaigns are integrated into a single estimate, making it challenging to isolate individual contributions and processes. This is especially problematic when the measurement is integrated over multiple years, and includes nearby disruptive events such as the 2015 summit intrusion that change the gravity landscape through the emplacement of mass and associated deformation. Seasonal and hydrological effects, such as rain and snow melting, may induce significant subsurface mass changes [*Miller et al.*, 2017; *Poland and de Zeeuw-van Dalfsen*, 2019; *Carbone et al.*, 2019] that are usually difficult to estimate accurately. At Kīlauea, the water table is situated roughly half a kilometer below the surface and experiences minor fluctuations [*Kauahikaua*, 1993; *Johnson et al.*, 2010]. The local gravity change from transient hydrological effects is therefore expected to be minor and not considered here.

Uncertainties in vertical deformation estimates are estimated to be about 2 cm ($\approx$ 6 µGal), representing the maximum mismatch between the GNSS and InSAR observations. Deviations from the theoretical free-air gradient impose additional uncertainties that are not easily quantified. The local free-air gradient depends significantly on the source of deformation and may be different for e.g., post-glacial rebound [*Olsson et al.*, 2015] compared to volcano deformation involving subsurface fluid redistribution, where the free-air gradient or Bouguer corrected free-air gradient [*Vajda et al.*, 2020, 2021] may be more suitable. Free-body geometry inversions [*Camacho et al.*, 2021] or coupled inversions of surface deformation and gravity [*Nikkhoo and Rivalta*, 2022] may provide an alternative to recover source estimates; however, mass accumulation without commensurate surface deformation that involves non-elastic behavior, e.g., density changes through degassing or the compressibility of gas-rich magma [*Rivalta and Segall*, 2008], makes the application of such models nontrivial. Furthermore, because multiple processes and sources may have been active over the 2009 – 2015 period, we adopted a classical approach, applying a (theoretical) correction for the observed vertical surface deformation before completing point source gravity inversions.

■ 4.4.3 Microgravity Changes during 2009 – 2017 at Kīlauea

Shallow Mass Accumulation without Commensurate Surface Deformation

Campaign microgravity results indicate that over the course of the 2008 – 2018 summit eruption, mass accumulated at shallow depth beneath Kīlauea Caldera (fig. 4.4 and table 4.2). The point source solution is the most robust when the microgravity observations are integrated over eight campaigns between 2009 and 2015, indicating mass accumulation on the order of 1.9×10^{11} kg at a depth of roughly 1.3 km below sea level. The source inversion from the April 2017 campaign yielded unrealistic results because of its sparse measurements and low data quality. The recovered depth and position of mass accumulation is slightly below the expected level of the HMMR at approximately 0 to 1 km below sea level [*Poland et al.*, 2014; *Bemelmans et al.*, 2021]. The discrepancy between the expected depth of the HMMR and the recovered center of mass accumulation may be explained by the trade-off in gravity between depth and mass, and the added challenge of recovering accurate depth estimates from surface measurements. Another probable cause is that the observed gravitational effect is a combination of mass accumulation in both the HMMR and SCR, integrated in a single estimate. Furthermore, the recovered depth represents a point source, and the *real* spheroid reservoir boundaries may overlap with geodetic sources from the literature. Alternatively, geodetic and gravity changes may be sensitive to other distinct parts of the plumbing system.

The relatively continuous subsurface mass accumulation occurred even during periods characterized by subsidence (fig. 4.4; *Poland et al.* [2021b]). Deformation at Kīlauea during the 2008 – 2018 eruption began as subsidence between 2009 – 2010, progressed towards minor uplift inside Kīlauea Caldera during 2011 – 2012, and turned into significant uplift at the estimated surface position of the SCR following May 2014. The rapid rate of uplift that occurred after May 2014 in the south part of Kīlauea Caldera marked a new phase in the eruption, with increased pressurization eventually leading to a magmatic intrusion in the summit area during May 2015 [*Jo et al.*, 2015; *Bemelmans et al.*, 2021]. Even during this period of intense inflation above the SCR, microgravity observations indicate that mass continued to accumulate mainly near the HMMR. This discrepancy may arise from the fact that mass accumulation in the SCR was too distant from the dense network of benchmarks to be fully characterized. Any such deep variations in mass may also be obscured by shallower mass accumulation near the HMMR. In case half of the observed mass accumulates near the HMMR at a reference depth of 500 m b.s.l. and the other half occurs in the SCR at 3000 m b.s.l. [*Poland et al.*, 2021b] the felt gravity change at the surface above the reservoirs would be 280 and 40 µGal respectively. The individual contributions are difficult to estimate because only the integrated amount of gravity change is observed at the surface.

The apparent discrepancy between microgravity observations and surface deformation has been recognised by all previous investigations that compared microgravity campaigns and surface displacements at Kīlauea [*Dzurisin et al.*, 1980; *Johnson et al.*, 2010; *Bagnardi et al.*, 2014]. Assuming a simple Mogi point source [*Mogi*, 1958] located at the recovered 1.3 km depth b.s.l., with a total volume change of $9.5 \times 10^7\,\mathrm{m}^3$ ($\delta V = \delta M/\rho$), calculated from the recovered 1.9×10^{11} source mass (δM), and an assumed magma density of $2000\,\mathrm{kg\,m^{-3}}$ (ρ). With these parameters, the expected vertical deformation at the surface above the source would exceed 4 m, while the observed vertical deformation (≈ 20 cm) explains only about 5 % of the predicted deformation. This simple estimate suggests that a process is required that accommodates mass increase without corresponding volume increase.

Dzurisin et al. [1980] and *Johnson et al.* [2010] proposed that the driving process may be the filling of voids below the surface of Kīlauea in a network of interconnected cracks. The existence of such void space is indicated by the difference between the volume of the summit eruptive vent and that of the associated ejecta following its formation in March 2008 – the ejecta accounted for only 1 % of the volume of the source crater [*Houghton et al.*, 2011]. However, part of the material may have been assimilated in the plumbing system and flowed towards the ERZ. Likewise, gravity changes measured during 2018 – 2019, following the summit collapse and lower East Rift Zone eruption, suggested the presence and filling of voids beneath Kīlauea Caldera [*Poland et al.*, 2019]. Complementary to the idea of void space, *Bagnardi et al.* [2014] offered an alternative explanation that invoked densification of magma in the reservoir through degassing or the compressibility of gas-rich magma [*Rivalta and Segall*, 2008] – in which the void space is effectively contained within the magma itself.

Lava Lake and Reservoir Density Changes

The gravity response to the summit lava level was greatest at three benchmarks: HOVL-G, HVO41, and 204YY, in order of proximity from the summit vent (HVO41 is illustrated fig. 4.1 and the benchmarks nearest to the lava lake are HOVL-G and 204YY respectively). *Bagnardi et al.* [2014] removed the effect of changes in the lava lake on these benchmarks using a geometric model of the conduit before completing point source inversions and found comparable point source estimates to the results presented here. Here, the point source inversions are completed excluding the observed changes at these benchmarks. Excluding these benchmarks reduces the biasing of the point source inversion results towards the eruptive vent, in case such a geometric forward model fails to capture the observed disruptive vent changes between campaigns. As an added benefit, the recovered point sources (table 4.2) can be used to isolate the gravity effect of the lava lake at these benchmarks (fig. 4.8, bottom). Comparison of results between the forward model used by *Bagnardi et al.* [2014] and that from fig. 4.8 (bottom) shows that the gravity effect of the lava lake appears similar. The isolated effect between March 2011 – November 2012 shows a similar increase for benchmarks HVO41 and 205YY (100 µGal) as calculated by the forward model of *Bagnardi et al.* [2014]. The isolated effect on benchmark HOVL-G however, is about 50 to 100 µGal higher, presumably because it is closer to the source and more sensitive to inaccuracies.

The period stretching from March 2011 to June 2012 spans the most variation in the lava lake level (fig. 4.8). The nearest benchmarks show an increase in gravity that is small compared to the much steeper increase in gravity between June 2012 and November 2012, despite a smaller rise in the level of the lava lake. This difference may imply that the density of the lava lake was higher in 2012 compared to 2011, producing larger gravity change for smaller lava lake level change. In the period between November 2012 and September 2015, the level of the lava lake remained relatively stable at an elevated level, except during the May 2015 summit intrusion [*Bemelmans et al.*, 2021]. The decrease in gravity at the benchmarks nearest to the summit eruptive vent between November 2012 and November 2013 may be a result of the rim collapse that occurred in January 2013 and the loss of 21 000 m^3 of material [*Patrick et al.*, 2019b], effectively replacing solid rock with air. After 2014, InSAR data (fig. 4.4 – row 1) show consistent inflation without commensurate changes to the level of the lava lake. In the case that the lava lake is in magmastatic equilibrium with the deeper magmatic system, an increase in subsurface pressure that leads to surface deformation would also probably lead to an increased lava level inside the summit vent. Such rise in the lava lake level is not observed, and the stability of the lava lake level over this period may instead be explained by an increase in magmastatic pressure inside the summit vent. The density increase may occur in the reservoir or at the top of the conduit and be recycled into the deeper reservoir after degassing. Such magmastatic increase in pressure as a result of higher magma density ($\rho g h$) would counteract the increased deeper pressurisation in order to keep the lava level stable, while the surrounding region continued to inflate. Alternatively, *Patrick et al.* [2019b] suggest that after 2014 the lava lake level does not change proportionally with deformation because the pressurisation occurs within the SCR

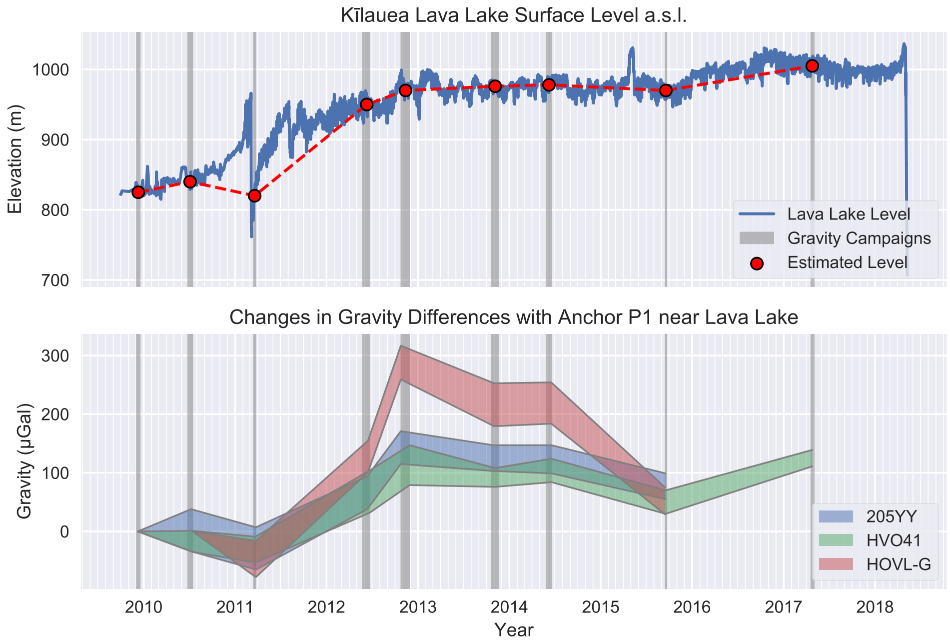

Figure 4.8: *Top) Absolute Kīlauea lava lake levels from sea level between 2010 and 2018 on Kīlauea, Hawai'i [Patrick et al., 2019b]. The time spans of the microgravity campaigns from fig. 4.2 are marked by vertical grey bars. Bottom) Observed changes in gravity difference of three benchmarks closest to the lava lake (HOVL-G, 205YY and HVO41) with confidence ranges. The data have been corrected for the contribution of the recovered point source solutions (table 4.2). Since 2011, the level has been steadily rising outside of a major disruptive event at Kamoamoa in the beginning of 2011. Beyond 2013 the level of the lake remained stable besides the 2015 summit intrusion, until ultimately draining completely after the outbreak in the lower East Rift Zone (ERZ) in 2018. a.s.l. = above sea level.*

instead of the HMMR (fig. 4.4). This hypothesis requires that the SCR and HMMR are not strongly connected. With a direct connection between the reservoirs, deep pressurisation would be felt throughout the entire plumbing system, including the HMMR, and be expressed in the lava lake level. The suggested hypothesis for the stability of the lava lake during this period thus appears contingent on the assumed configuration of the plumbing system. After September 2015, the level of the lava lake rose steadily, but changes in microgravity for this period were only available for benchmark HVO41. At that time, it may be that the density of the magma was reaching its bubble-free limit – forcing pressure increases coming from the deeper magmatic system to be accommodated by a rise in the lava lake level once more.

A quantitative analysis of microgravity results from the summit vent benchmarks is challenging because of the long interval between campaigns, and the fact that major disruptive events modified the vent geometry. Continuous gravity observations from the rim of the summit eruptive vent support that lava lake density did increase over time, from roughly $950 \pm 300\,\mathrm{kg\,m^{-3}}$ in 2011 [*Carbone et al.*, 2013], climbing towards 1000 to $1500\,\mathrm{kg\,m^{-3}}$ between 2011 and 2015 [*Poland and Carbone*, 2016], and up to $1700\,\mathrm{kg\,m^{-3}}$ by the time of the 2018 lower ERZ eruption [*Poland et al.*, 2021a].

An increase in magma density in the lava lake inside the summit eruptive vent may indicate, when considering the observation of subsurface mass accumulation without commensurate surface deformation, that densification also took place throughout the subsurface magma reservoirs. Significant variations in magma density in the shallow reservoir have also been detected through seismic techniques [*Crozier and Karlstrom*, 2021], and significant degassing took place based on the increased rate of SO_2 emissions coming from the summit [*Sutton and Elias*, 2014; *Elias et al.*, 2018] after the opening of the vent. Similarly, the frequency of gas pistoning events – sudden changes in lava level driven by gas accumulation and release in the upper part of the lava lake – decreased from 2010 – 2015 [*Patrick et al.*, 2016], indicating that less gas-rich, denser magma was becoming more prevalent in the reservoir. Multiple independent observations, including the campaign microgravity results presented here, thus support the hypothesis of densification of magma within the magmatic system.

Implications of Density Changes for Magma Supply Rates

The magma supply rate at Kīlauea has a fundamental impact on the character (e.g., rate, volume, and duration) of its eruptions [*Swanson*, 1972; *Dvorak and Dzurisin*, 1993; *Poland et al.*, 2012]. Variations in the supply rate of Kīlauea's magmatic system have previously been estimated directly from outflow volumes, or through proxy measurements such as ground deformation or gas emissions. These estimates indicated an above average supply rate between 2003 – 2006 before the summit eruption began, and a potential lull in magma supply rate between 2011 and 2012 [*Anderson and Poland*, 2016], likely returning to pre-eruption levels by 2016 [*Dzurisin and Poland*, 2019]. The magma-supply models are often based on mass-balance equations that assume incompressible magma and thus do not account for the potential influence of magma density changes inside the magma storage reservoirs. One consequence of this approach is that the inferred decrease in supply during 2011 – 2012 may potentially be overestimated because (some portion of) the supply was accommodated by the compression and densification of gas-rich magma. The presented estimate of the amount of subsurface mass accumulated beneath Kīlauea Caldera from microgravity campaigns may provide additional constraints that might improve magma supply rate estimates.

Potential Effects of Densification on Eruption Behaviour

Another implication of magma densification is that a column of high density magma may increase pressure in other parts of the magmatic system. One possible example is the summit intrusion in May 2015, which was preceded by a rise in lava level overflowing the summit vent in Halemaʻumaʻu crater. The sudden rise in lava lake level was followed by a drop of similar magnitude, as an intrusion was emplaced below the southern part of Kīlauea Caldera [*Patrick et al.*, 2019b; *Bemelmans et al.*, 2021], providing an alternative outlet for the excess pressure. A similar mechanism may have contributed to expediting the lower ERZ eruption in 2018. Increased (magmastatic) pressure from Kīlauea's summit area was probably felt throughout the volcano's magmatic system, as indicated by inflation that was especially strong along the ERZ and at the summit in the weeks prior to the 2018 lower ERZ outbreak [*Patrick et al.*, 2020; *Poland et al.*, 2022]. The increased magmastatic pressure of the magma filling Kilauea's summit reservoirs may thus have contributed to the breaking of a barrier towards the ERZ on April 30, which allowed the summit reservoir and lower-elevation plumbing system of dense accumulated magma to feed the LERZ eruption, as is expected from microstructural constraints of the erupted magma [*Wieser et al.*, 2020]. Microgravity results can provide no additional information on the configuration of the plumbing system below Kīlauea Caldera. Whether the ERZ is sourced by the HMMR [*Wang et al.*, 2021] or the SCR [*Wieser et al.*, 2020] is unknown.

4.5 Conclusion

Nine microgravity campaigns were completed at Kīlauea between 2009 – 2017. The data were reduced using a weighted least-squares approach [*Hwang et al.*, 2002], which proved especially effective in short double-loop circuits. Derived microgravity changes over time illustrate that subsurface mass on the order of 1.9×10^{11} kg accumulated beneath Kīlauea Caldera during 2009 – 2015. The mass increase was identified at an average depth of 1.3 km below sea level, slightly below the expected position of the shallow Halemaʻumaʻu reservoir from alternative geodetic observations. The accumulation of mass had probably been occurring since 1975 [*Johnson et al.*, 2010; *Bagnardi et al.*, 2014] and continued until the 2018 eruptive outbreak in the lower East Rift Zone. The accumulation of subsurface mass beneath Kīlauea Caldera was not commensurate with the observed surface deformation. This discrepancy may be caused by densification of magma inside the Halemaʻumaʻu reservoir through a combination of degassing through the summit eruptive vent and the densification and compression of gas-rich magma [*Rivalta and Segall*, 2008] – a conclusion that might impact apparent changes in magma supply over time and that argues for the inclusion of microgravity data in modeling of magma supply rates. When magma compressibility and densification continue to counteract increasing pressure from the deeper plumbing system, excess pressure may be relieved through alternative means. Such transfer of pressure accommodation mechanisms in Kīlauea's summit area might explain activity like the 2015 summit intrusion and may have even expedited the devastating lower East Rift Zone outbreak in 2018.

4.6 Acknowledgments

We would like to thank the JGR Solid Earth (associate) editor for their comments. Reviewers Taiyi Wang, Craig Miller, and Kyle Anderson are greatly thanked for their thorough reviews of the manuscript. We also express our gratitude to the USGS staff for their suggestions. Your comments have led to the serious improvement of many aspects of this paper. Maurizio Battaglia is thanked for discussion and suggestions on how to apply ocean loading corrections and Duncan Agnew is thanked for his advise using HARDISP. We are grateful to the staff of the Hawaiian Volcano Observatory for their support of gravity measurements at Kīlauea. All of the data described in this report were collected within Hawai'i Volcanoes National Park, and we are grateful for the collaboration and support of numerous Hawai'i Volcanoes National Park staff over the years in facilitating our gravity monitoring and research efforts. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This work was supported by the NEWTON-g project, which has received funding from the EC's Horizon 2020 programme, under the FETOPEN-2016/2017 call (Grant Agreement No 801221).

5

Decades of Subsidence Followed By Rapid Uplift: Insights from Microgravity Data at Askja Volcano, Iceland

Abstract In August 2021, Askja volcano in Iceland returned to the spotlight after a sudden onset of rapid uplift followed decades of continuous subsidence. In this study the extended record of microgravity data from Askja between 1988 to 2017 is revisited, and new microgravity data from 2021 and 2022 are introduced, which were collected after the uplift had started. Askja caldera had been steadily subsiding since at least 1984 and was characterised by a net decrease in microgravity, potentially signalling the contraction of its magma chamber or eviction of magma either laterally or to deeper levels. The microgravity data indicate that despite ongoing subsidence between 2017 and early 2021, a significant gravity increase can be detected in the center of the caldera between 2017 and August 2021. This increase may be introduced during – or leading up to – the period of uplift. The new microgravity data also indicate that during the period of 40 cm uplift after August 2021 to fall 2022, gravity changes approach the free-air gradient, suggesting subsurface density decreases as a driving process. This process may relate to the vesiculation of magma previously emplaced in the volcano roots, a change in the hydrothermal system, or replacement of dense basaltic magma with less dense rhyolitic magma, or a combination of these processes. However, uncertainties for this period are elevated and may obscure a gravity signal expected from additional mass accumulation. The timing and high uncertainties of some campaigns make it challenging to be conclusive on the driving process behind the uplift, but future microgravity campaigns

Published as: Koymans et al., (2023). Decades of subsidence followed by rapid uplift: Insights from microgravity data at Askja Volcano, Iceland, Journal of Volcanology and Geothermal Research, Volume 442, 10789, doi: 10.1016/j.jvolgeores.2023.107890

could help solve the ambiguity. The study also provides a description of potential pitfalls in microgravity campaigns and recommendations on how the reliability of microgravity data can be improved.

5.1 Introduction

Long-term geodetic monitoring requires dedication. Worldwide, only a handful of volcanoes (e.g., Campi Flegrei [*Berrino*, 1994; *Gottsmann et al.*, 2003], Kīlauea [*Johnson et al.*, 2010; *Bagnardi et al.*, 2014; *Koymans et al.*, 2022], and Krafla [*Rymer et al.*, 1998]) have such a uniquely extensive deformation and microgravity record as Askja, Iceland. The remarkable geodetic record of Askja enables the study of the temporal evolution of the volcano, and covers a sudden reversal in August 2021 from a four decade-long interval of subsidence of more than 1 m, that changed to rapid uplift at a rate of up to 40 cm per year. It presents an opportunity to study the geodetic signatures and to identify the source that is responsible for the observed change, and is key for hazard implications. In this pursuit, microgravity observations are beneficial because it is the only technique that can identify any potential change in subsurface mass below the caldera. One key objective of microgravity surveys in volcano monitoring is thus to detect gravimetric signatures that may indicate, or represent precursors to major changes in the character of volcanic activity that would otherwise remain undetected (e.g., *Rymer* [1994]; *Battaglia et al.* [2008]; *Poland et al.* [2021a]).

Askja is located in the desolate highlands of central Iceland and lies at the heart of the Northern Volcanic Zone (NVZ), at the divergent boundary between the Eurasian and North American plates (fig. 5.1). Near Askja, the NVZ is oriented north-south and extends in length over more than 150 km, and is locally up to 10 to 15 km wide [*Sigvaldason*, 1979]. The divergent plate boundary hosts multiple active volcanic systems, including Askja, Bárdarbunga, Grimsvötn, and Krafla that accommodate the strain produced by the plate spreading [*Drouin et al.*, 2017]. Askja consists of nested calderas, where the main caldera has a diameter of 7 to 8 km (labeled Askja; fig. 5.1), and potentially formed as the result of a Plinian eruption during the early Holocene. Alternatively, the rim of the main caldera may have been built up through sub-glacial fissure eruptions [*Brown et al.*, 1991], accompanied by gradual caldera collapse [*Gudmundsson et al.*, 2016]. The smallest and youngest caldera (labeled Öskjuvatn; fig. 5.1) developed as the result of an explosive eruption in 1875 [*Sigvaldason*, 1979; *Sparks et al.*, 1981; *Hartley and Thordarson*, 2012]. This caldera presently hosts one of the deepest lakes in Iceland. The most recent eruption at Askja was effusive and dates back to 1961 [*Thorarinsson and Sigvaldason*, 1962]. During this eruption, basaltic lava flowed through the Öskjuop pass in the northeast of the caldera onto the flanks of the volcano, creating a convenient path into the caldera.

Askja demonstrates an intermediate level of seismicity that is associated with hydrothermal activity [*Greenfield et al.*, 2020; *Winder*, 2022]. Hydrothermal activity is extensive and partly focused on caldera rims, but also identified surrounding lake Öskjuvatn [*Ranta et al.*, 2023], and in the lukewarm muddy waters of Víti, which can comfort only the most intrepid tourists.

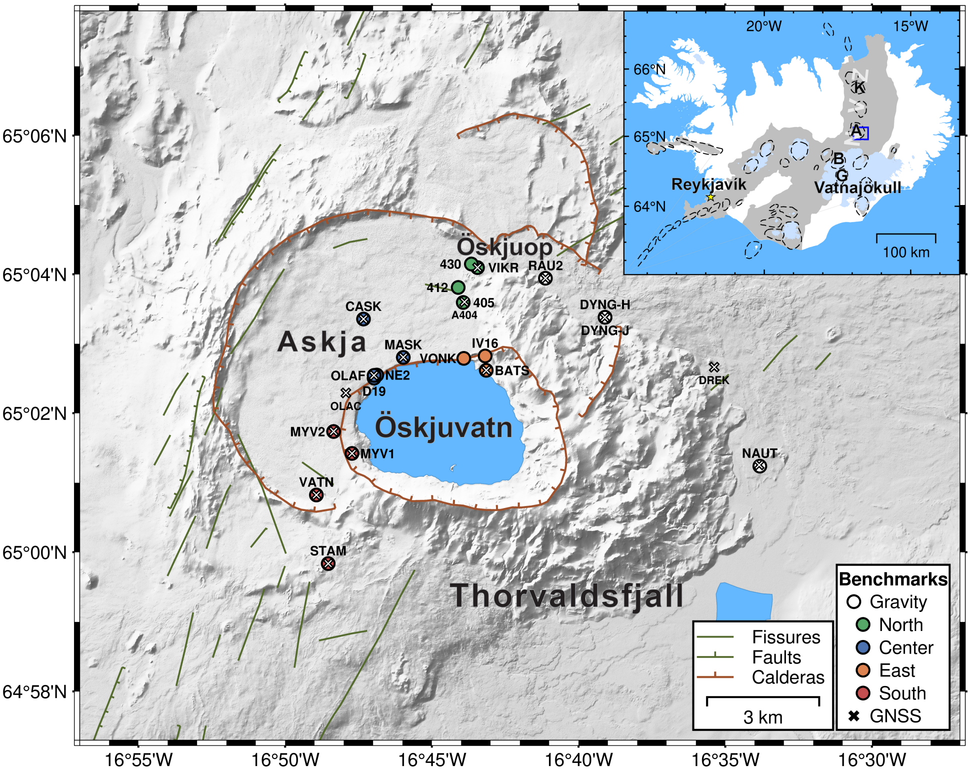

Figure 5.1: *Hillshaded digital elevation model of Askja highlighting the location of the caldera rims in brown [Hartley and Thordarson, 2012]. The map inset shows the location of Askja and illustrates glaciers (light blue) and Icelandic volcanic zones (grey). The Northern Volcanic Zone (NVZ) stretches from the north of Krafla (K) to Askja (A) and towards Bárðarbunga (B) and Grímsvötn (G). The microgravity network at Askja consists of 20 microgravity benchmarks with their identifiers. The colors represent different regional groups of microgravity benchmarks. The crosses indicate locations with continuous GNSS receivers or campaign benchmarks.*

5.1.1 Geodetic Monitoring at Askja

After the 1961 effusive eruption, a monitoring network was designed, targeted to capture the volcano's temporal evolution with a diverse set of geodetic tools (fig. 5.2). Precise levelling data from a 1.7 km line are available from 1966 to 1972 [*Tryggvason*, 1989], and the line has been measured annually since 1983 [*Sturkell and Sigmundsson*, 2000; *Sturkell et al.*, 2006; *de Zeeuw-van Dalfsen et al.*, 2013]. In the 1990s, campaign (since 1993) and continuous (since 2000) Global Navigation Satellite System (GNSS) measurements were started, and Interferometric Synthetic-Aperture Radar (InSAR) observations (since 1992) began to be included in volcano geodesy.

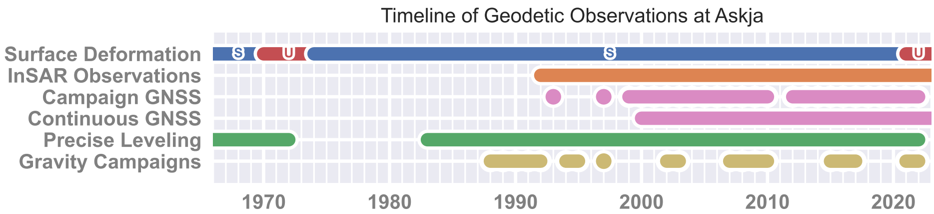

Figure 5.2: *Approximate coverage of geodetic monitoring techniques during periods of uplift and subsidence (U – Uplift, S – Subsidence). A continuous bar for campaign measurements indicates that data are available at least once per year. See table 5.1 for the precise micro-gravity campaign dates and fig. 5.3 for the available campaign GNSS data since 2012.*

5.1.2 Subsurface Structure and Evolution of Askja

Leveling data indicate uplift of Askja caldera between 1970 and 1973 [*Tryggvason*, 1989; *Sturkell et al.*, 2006], followed by an extended 40 year period of slowly decaying subsidence since at least 1984, and potentially as far back as 1974 when interpolated between measurements [*Sturkell et al.*, 2006]. The character of this subsidence was that of stable exponential decay with an estimated relaxation time of 39 – 42 years [*Sturkell et al.*, 2006; *Giniaux*, 2019], and an inferred total subsidence in the center caldera of over 1 m. Deformation data from decades of subsidence indicate that a shallow magma body is likely located at approximately 2 to 3 km depth [*Pagli et al.*, 2006; *Sturkell et al.*, 2006; *de Zeeuw-van Dalfsen et al.*, 2013]. Nearly all modeled geodetic sources indicate a deflating point pressure source [*Mogi*, 1958] at approximately this depth below the center of the main caldera [*Tryggvason*, 1989; *Rymer and Tryggvason*, 1993; *Sturkell and Sigmundsson*, 2000; *de Zeeuw-van Dalfsen et al.*, 2012; *Drouin et al.*, 2017]. An elliptical source model [*Pagli et al.*, 2006], and two distinctive Mogi sources [*Sturkell et al.*, 2006] – at depths of 3 km and 16 km, respectively – have been proposed as alternatives. Seismic tomography reveals features that represent a shallow magma storage area at 5 to 6 km depth b.s.l., and the potential existence of a magma mush storage and transport zone at 10 to 25 km depth [*Mitchell et al.*, 2013]. These models do not have sufficient resolution at 2 to 3 km and thus do not rule out the potential presence of magma at shallow depths. However, a shallow ($\leq$ 3 km b.s.l.) high seismic velocity zone below the caldera may indicate an intrusive complex, with an observed low V_p/V_s ratio also suggesting the phase transition from water to steam at this depth [*Halldórsdóttir et al.*, 2010; *Greenfield et al.*, 2016].

In August 2021, the decades-long trend of subsidence reversed, and the center caldera floor began to rise at a rate of up to 40 cm per year. The present rate of uplift identified from leveling observations indicate similar rates that were derived from leveling observations between 1971 – 1973 [*Sturkell et al.*, 2006].

Seismic tomography and deformation modeling provides insight into source volumes, or their changes and locations, but cannot exclusively determine which mechanism is responsible for the observed surface deformation of the caldera. Microgravity surveys add information on subsurface mass changes to bridge this observational gap, and together with surface deformation data can better constrain the governing volcanic processes. Microgravity observations were started at Askja in 1988 and completed episodically in the following decades up until 2022, providing a total of 19 microgravity surveys (table 5.1). For the extended period of subsidence, different source mechanisms have been suggested, such as a cooling and contracting magma chamber, and the flow from a shallow magma body to deeper levels, or through lateral movement [*de Zeeuw-van Dalfsen et al.*, 2005]. The stage of uplift that started during August 2021 has not yet been thoroughly studied and its cause remains enigmatic, but the inflection point from subsidence to uplift has been captured by geodetic observations.

In this study the full microgravity record of Askja is evaluated, joining new and previously published data. In the following sections the existing historical microgravity observations between 1988 – 2010 are reviewed, data from 2015 – 2017 are re-analysed based on available raw data, and two new microgravity campaigns from 2021 and 2022 are presented. The scope of this work focuses on the period since 1988 leading up to inflection point from extended subsidence to uplift. Microgravimetric signatures associated with the observed long-term trends in deformation are studied to investigate subsurface changes that may otherwise have remained undetected, and shed light on the governing magmatic and hydrothermal processes at Askja.

5.2 Methodology

5.2.1 Microgravity Campaigns at Askja

Over the past three decades, microgravity campaigns were also completed by various institutes and operators using state of the art equipment of their time (table 5.1). Microgravity results between 1988 – 1991 were collected by *Rymer and Tryggvason* [1993] using two LaCoste & Romberg (L&R) model G instruments. Additional surveys were completed between 1992 and 2003, using similar L&R model G gravimeters, extending the established record [*de Zeeuw-van Dalfsen et al.*, 2005]. These authors also improved previously estimated net gravity changes between 1988 – 1991, with the latest two-point Mogi source model at the time to correct for height changes [*Sturkell et al.*, 2006]. Between 2007 – 2009, microgravity data were collected by *Rymer et al.* [2010], including another campaign completed in 2010 [*de Zeeuw-van Dalfsen et al.*, 2013]. Campaigns were restarted by *Giniaux* [2019], who added three surveys between 2015 – 2017 using a Scintrex CG-5 gravimeter. Microgravity data were also collected in September 2021 and August 2022 in response to the observed uplift at Askja. These surveys were completed using two pairs of Scintrex CG-5, and a Scintrex CG-5 and CG-6, respectively.

Table 5.1: Microgravity campaigns at Askja, Iceland between 1988 – 2022. [1] Rymer and Tryggvason [1993], [2] de Zeeuw-van Dalfsen et al. [2005], [3] Rymer et al. [2010], [4] de Zeeuw-van Dalfsen et al. [2013], [5] Giniaux [2019], [6] this study. The Mogi source model for the vertical deformation correction in the 1988 – 2010 campaigns was published by Sturkell et al. [2006]

Year	Instruments and Serial Numbers	Deformation Correction	Ref
1988	L&R model G (513 and 105)	Mogi source model	[1, 2, 4]
1989	L&R model G (513 and 105)	Mogi source model	[1, 2, 4]
1990	L&R model G (513 and 105)	Mogi source model	[1, 2, 4]
1991	L&R model G (513 and 105)	Mogi source model	[1, 2, 4]
1992	L&R model G (513)	Mogi source model	[2, 4]
1994	L&R model G (513)	Mogi source model	[2, 4]
1995	L&R model G (513)	Mogi source model	[2, 4]
1997	L&R model G (513 and 403)	Mogi source model	[2, 4]
2002	L&R model G (513 and 403)	Mogi source model	[2, 4]
2003	L&R model G (513 and 403)	Mogi source model	[2, 4]
2007	L&R model G (513)	Mogi source model	[3, 4]
2008	L&R model G (513)	Mogi source model	[3, 4]
2009	L&R model G (513)	Mogi source model	[3, 4]
2010	L&R model G (513)	Mogi source model	[4]
2015	Scintrex CG-5 (968)	GNSS Measurements	[5]
2016	Scintrex CG-5 (968)	GNSS Measurements	[5]
2017	Scintrex CG-5 (968)	GNSS Measurements	[5]
2021	Scintrex CG-5 (41301), CG-5 (41421)	GNSS Measurements	[6]
2022	Scintrex CG-5 (41301), CG-6 (19090203)	GNSS Measurements	[6]

5.2.2 Askja Microgravity Network

The microgravity network at Askja has evolved over the past decades and consists of twenty benchmarks (fig. 5.1). The network of benchmarks was initially designed and set up by *Rymer and Tryggvason* [1993], extended by *de Zeeuw-van Dalfsen et al.* [2005] in 2002 with the addition of NE2, MASK, 430, and DYNG-H, and further developed by *Giniaux* [2019] in 2015 through the introduction of benchmarks CASK, DYNG-J, MYV1, MYV2, STAM, and VATN to improve the spatial coverage particularly in the south of the caldera. The highest density of benchmarks (MASK, OLAF, D19, and NE2) is in the center of the caldera, near the center of past subsidence and observed uplift since August 2021. After the campaign in 2022, benchmark VONK was marked for removal from the network because it is not mounted on a solid foundation and was previously measured in soil. Similarly, benchmark IV16 sits atop of a loose boulder in a region that is subject to heavy erosion and difficult to measure consistently. During the 2022 campaign, benchmark

430 appeared extremely unstable, both during the microgravity survey and the leveling measurements. Benchmark NAUT was established for future surveys at a stable point far away from the deforming caldera at an existing benchmark used for campaign GNSS measurements.

Microgravity Network Anchors

During the microgravity campaigns between 1988 and 2010, benchmark VIKR was selected as the network anchor. Beyond 2002, it was recognised that a new benchmark at DYNG, which eventually became co-located with a continuous GNSS receiver, provided a more suitable setting outside the area of active subsidence. At this site, two anchors were established: One anchor (DYNG-H) is located just beside the GNSS receiver, with the second anchor (DYNG-J) at 7 m distance. In the campaigns of 2015, 2016, and 2017, *Giniaux* [2019] measured gravity differences relative to DYNG-J, while other campaigns in the past included an occasional measurement of DYNG-H. Both benchmarks DYNG-H and DYNG-J were measured in the 2021 and 2022 campaigns, and the effective gravity difference between the two anchors is well-constrained at a positive 170 to 180 µGal going from DYNG-H to DYNG-J. For continuity with the historical time series, all microgravity data presented in this study, including those from recent surveys (2015 – 2022) for which measurements at DYNG are available, were expressed relative to the historical anchor VIKR. Since VIKR is located in the region of deformation it may be sensitive to subsurface mass changes and net microgravity changes may be underestimated. It is recommended that the gravity network anchors are measured relative to (preferably) an absolute reference point far outside of the caldera to identify such potential problems, but logistically this may be unfeasible.

Microgravity differences between the network anchor and the benchmarks from the 1988 to 2010 campaigns were taken from the literature [*Rymer and Tryggvason*, 1993; *de Zeeuw-van Dalfsen et al.*, 2005, 2013]. Data collected by *Giniaux* [2019] and the campaigns of 2021 and 2022 were (re-)analysed using the online tool presented in *Koymans et al.* [2022], simultaneously solving for instrumental drift and gravity differences (e.g., *Hwang et al.* [2002]), with an independent linear drift function fitted to each campaign day. Due to a limited number of measurements in the 2017 campaign, instrumental drift was fitted over the full campaign instead of daily. The microgravity data were corrected for the effect of the solid Earth tides using the default applied Scintrex tide correction [*Longman*, 1959]. An ocean loading and polar motion correction may provide only an insignificant 1 to 2 µGal improvement to the results and was not applied.

■ 5.2.3 Microgravity Campaigns of 2021 and 2022

Besides reviewing and partially re-analysing the existing record of microgravity data (1988 – 2017), this study presents new data that were collected in field campaigns during September 2021 and August 2022. Both campaigns were completed with two instruments in a sub-loop (A → B → C → B → C → A), mirrored form (A → B → C → C → B → A), or a modified version thereof. Each benchmark was occupied at least twice during a campaign day, completing at minimum two sets of 5 min measurements per occupation per instrument. This strategy provides a robust dataset with a sufficient number of repeated measurements. The campaign of September 2021 is considered unreliable because it was completed in poor weather conditions, including the occasional snowstorm. In particular, high wind speeds produce high variance in the microgravity data that can be recognised in the scatter of the measurements with uncertainties of up to 20 to 30 µGal on the windiest days. The August 2022 campaign was completed in much better weather conditions and produced robust readings with an associated uncertainty on the mean gravity difference with the network anchor of 2 to 3 µGal.

5.2.4 Deformation Correction

Corrections for relative changes in height with respect to the network anchor are made to the microgravity data. This correction is completed using the theoretical free-air gradient (FAG) of 308 µGal/m to isolate the effect of subsurface variations in mass on the microgravity observations.

Microgravity data between 1988 – 2010 are available from a series of previous publications (see table 5.1). Due to the lack of GNSS coverage at this time, these data were corrected for deformation using modeled subsidence from the double Mogi point source model proposed by *Sturkell et al.* [2006]. The change in vertical deformation between 2010 – 2015 was estimated by extrapolation from the GNSS data going back to 2012. For the campaigns between 2015 – 2022, height changes between the campaign dates were precisely estimated using co-located or measurements at nearby GNSS stations (see fig. 5.1). They are all expressed relative to the vertical displacement experienced at network anchor VIKR to cancel the deformation effect at the anchor.

Continuous and campaign GNSS data in Askja from 2012 to 2022 were processed at University of Iceland using GAMIT/GLOBK 10.75 [*Herring et al.*, 2010] together with over 100 globally distributed reference stations and continuous GNSS stations in Iceland. The solid Earth tide, ocean tide, and pole tide were corrected. The IGS final orbit products, ocean tide model FES2004 [*Lyard et al.*, 2006], and IGS ionosphere products were applied. Only Global Positioning System (GPS) signals were used to derive the coordinates of the benchmarks. The coordinates were derived in the IGb14 reference frame aligned with ITRF2014 [*Altamimi et al.*, 2016], and then converted to topocentric coordinate system (North, East, and Up).

5.3 Results

In the following sections, results from vertical deformation estimates are given, after which the net microgravity series from 1988 – 2022 expressed relative to VIKR are presented. In the following analysis, the microgravity benchmarks are grouped by region, following the same color scheme used in fig. 5.1. The microgravity changes are not plotted on a map because there appears to be no spatial coherence outside of the regional groups, and the results lend themselves better to be presented in the

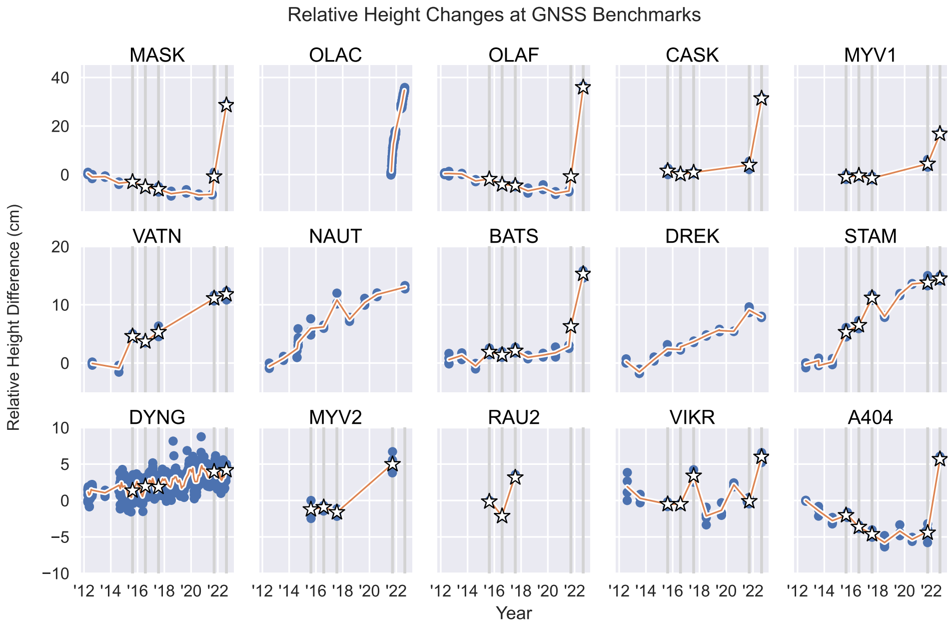

Figure 5.3: Inferred vertical surface displacement (cm) between 2012–2022, expressed relative to the first measurement at each GNSS benchmark (fig. 5.1). The blue dots represent campaign or continuous measurements and the orange curves interpolate between them. The vertical bars and white stars represent the selected height observations during the microgravity campaigns (table 5.1) that are used for the vertical deformation correction (excluding OLAC, NAUT, and DREK). Note the different y-axis scale that is the same for all panels in each row.

form of a time series graph.

5.3.1 Vertical Deformation Estimates

The geodetic network at Askja is well designed and favorable for microgravity analysis. Since GNSS benchmarks are co-located with (or nearly adjacent to) microgravity benchmarks (fig. 5.1), changes in height can be determined to within approximately 1 to 2 cm precision, equivalent to 3 to 6 µGal. In microgravity analysis, only the vertical component is considered to correct for the effect of the free-air gradient. Horizontal displacements experienced by benchmarks may be of significant amplitude, but will have no detectable effect on the microgravity measurements.

fig. 5.3 illustrates the vertical surface deformation observed at various GNSS benchmarks within Askja caldera (see fig. 5.1). Since 2012, GNSS benchmarks near the center of the caldera (e.g., OLAF and MASK) have been subsiding at a slow and steady rate of 1 to 2 cm per year. The reversal from subsidence to uplift of up to 40 cm per year in August 2021 can be clearly observed in benchmarks near the center of the caldera (OLAF, OLAC, MASK, and CASK). The uplift is also detected, although to a lesser extent, in the east (BATS-JH04*), south (MYV1), and northern part of the caldera (A404 and VIKR). These data clearly indicate that network anchor DYNG is a more suitable network anchor than VIKR for microgravity measurements, considering the small amount of vertical deformation between campaigns that is observed at DYNG outside of seasonal variations (1 to 2 cm). However, because historical data cannot be expressed relative to anchor DYNG, benchmark VIKR, that experiences some vertical deformation during inflation, had to suffice. Regardless of this suboptimal choice, the relative changes in height between the benchmarks and VIKR is naturally accounted for.

It is noteworthy that the signal that is being recorded during the period of uplift appears to be sensed in a wider geographic area by benchmarks that were not as strongly affected by the deflationary source (e.g., BATS-JH04).

5.3.2 Microgravity Series 1988 – 2022

fig. 5.4 shows net microgravity results (corrected for the free-air gradient) since 1988, expressed relative to network anchor VIKR and the 2022 measurement. The year 2022 was selected as the base year for comparison because all benchmarks were occupied during that campaign. Microgravity changes in the center of the caldera (blue) indicate a decreasing trend relative to the first measurement in 1988, while results from the north and east region (green and orange, respectively) scatter around zero observable change. The decreasing long-term trend in the center of the caldera reaches a minimum of about −150 to −170 µGal towards 2016 and appears consistently in the center benchmarks. A small potential diversion from this downward trend occurred in 2007 – 2008 (40 to 60 µGal) that was recognised by *Rymer et al.* [2010]. The possible causes that have been suggested are magma accumulation below the caldera despite continued subsidence, or changes in the hydrothermal system [*de Zeeuw-van Dalfsen et al.*, 2013]. The increase in gravity in the center benchmarks between 2016 and 2017 is challenging to verify, as the single center caldera measurement of 2016 may represent an underestimation. In 2021, a gravity increase (80 to 120 µGal) was detected relative to preceding years that cannot be accurately timed due to the extended data gap since the previous reliable campaign in 2016. During the period of uplift in 2021 and 2022, gravity did not change significantly in the center of the caldera relative to benchmark VIKR. Between 2015 – 2022 no significant changes are detected in the north and east benchmarks. It should be emphasized that these results are expressed relative to benchmark VIKR inside the caldera, which is undergoing active deformation and thus likely represent a lower bound on the gravity change estimate. Uncertainties of the historical gravity record reported in the literature are ±20 µGal [*de Zeeuw-van Dalfsen et al.*, 2005, 2013], and long-term trends should therefore be considered qualitatively.

The change from 2021 to 2022 in the center benchmarks may not be considered

*Updated Aug 2024 - in 2022 and 2023 benchmark BATS-JH04 was accidentally measured instead of BATS.

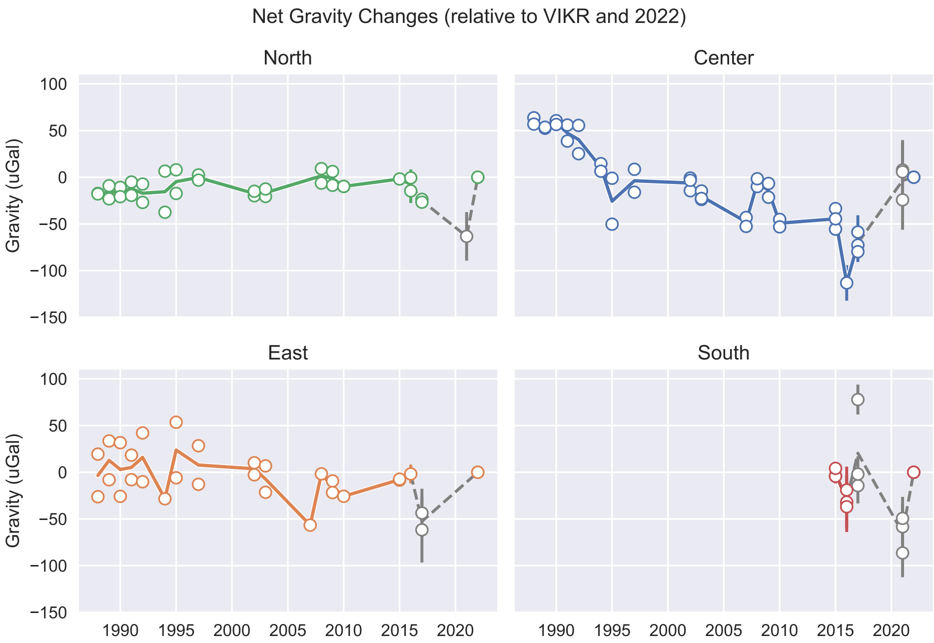

Figure 5.4: *Compiled net relative microgravity measurements expressed relative to anchor VIKR and 2022. The colors present regional groups illustrated in fig. 5.1. Microgravity changes were corrected for vertical deformation estimates using the theoretical free-air gradient of* 308 µGal/m. *Campaigns and groups characterised by the hybrid drift correction approach (2017) or large uncertainties (2021) are colored grey.*

significant because of the poor data quality of the 2021 campaign. Uncertainties in microgravity differences for the 2015 and 2016 campaigns vary between 10 to 20 µGal. The data from the east and south group of stations in the 2017 campaign have uncertainties on the order of 40 to 75 µGal. The reason for these elevated uncertainties is that a constant linear instrumental drift had to be fit over the entire campaign to determine gravity differences for these groups. This constraint was imposed by varying circuit anchors on different days and poorly tied measurements between the anchors. Fortunately, this approach could be avoided for the center and north benchmarks because they were in a direct circuit with anchor VIKR, providing sub-20 µGal confidence limits. The uncertainty of the mean gravity differences recovered in 2021 are on the order of 10 to 30 µGal, mainly due to the strong winds. Note that the reported uncertainties represent the inferred precision of the measurements but do not provide information on the accuracy, nor the repeatability of the measurement, particularly for benchmarks that were only occupied once in a campaign. The 2022 campaign has the highest data quality with uncertainties of less than 5 µGal. In this campaign, the benchmarks nearest to the observed uplift were also measured twice during two circuits on different days. These independent measurements of the central caldera benchmarks agree to the 1 to 2 µGal level – providing the desired confidence in these results. Besides the reported precision, unquantified systemic uncertainties in microgravity results and how they can be avoided are discussed in detail below.

5.4 Discussion

In the following section, the sensitivity of microgravity measurements is discussed with respect to campaign strategy, user preferences in processing, and the applied deformation correction. Furthermore, the possible unknown errors that remain in the microgravity data are highlighted, along with a discussion on the repeatability of microgravity measurements. The results from the microgravity time series are then discussed in relation to the observed caldera deformation patterns.

5.4.1 Sensitivity of Microgravity Measurements

Campaign microgravity measurements are sensitive to any type of mass redistribution [*Van Camp et al.*, 2017] and challenging to apply effectively due to their limited spatio-temporal resolution and inherently high uncertainties. The gravity effect of the level of lake Öskjuvatn, fluctuations in groundwater levels, atmospheric pressure, and snow variations cannot be accurately determined and have been neglected during the analysis, introducing an error of unknown amplitude to the microgravity results. However, all campaigns are completed in summertime when the caldera is (more or less) snow-free, limiting any potential short-period aliasing effects expected from different seasons. Besides, these effects have been shown to introduce an integrated effect that is limited to 20 µGal [*Giniaux*, 2019; *Poland and de Zeeuw-van Dalfsen*, 2019], within conventional demonstrated uncertainties in microgravity surveys (e.g., *Rymer* [1994]; *Battaglia et al.* [2003]; *Gottsmann et al.* [2003]; *Poland and de Zeeuw-van Dalfsen* [2019]; *Koymans et al.* [2022]). In addition to the uncertainty on vertical deformation measurements (1 to 2 cm), the applied gravity gradient also has an associated uncertainty. Measured free-air gradient estimates in the caldera vary between −240 to −360 µGal/m [*Rymer and Tryggvason*, 1993; *de Zeeuw-van Dalfsen et al.*, 2005]. As gradient measurements are not available for each benchmark and this value may change over time, the theoretical gradient is used for the deformation reduction of the gravity data. Furthermore, the experienced gravity gradient also depends on the character of the source (e.g., depth). A correction using the theoretical gradient is often preferred when the source responsible for the deformation is at least a few km deep [*Giniaux*, 2019]. Based on the measurements, the maximum additional error on the gravity change introduced by applying an inaccurate vertical gradient would approach −50 to 70 µGal/m (negative means that the gravity change was overestimated).

The microgravity campaigns studied in this study were completed with various types of instruments. The choice of instrument should in theory have no effect on the results because measurements are expressed relative to another point measured with the same instrument. Modern equipment (e.g., Scintrex CG-5, CG-6) is more convenient to operate, enabling multiple measurements in less time, assuring a more realistic drift correction. Newer gravimeters are also less prone to sudden measurement offsets (data tares). Gravimeter calibration factors may change over time (e.g., *Battaglia et al.* [2018]), but such an effect should in this case not influence the results significantly, because there is little variation in terrain elevation between the benchmarks at Askja, with a maximum difference of 200 m.

For the compiled microgravity record (1988 – 2022), it should be emphasized that observed gravity changes may be underestimated when they are expressed relative to network anchor VIKR located within the deforming caldera (figs. 5.1 and 5.3). Any mass variations strong enough to be sensed at the microgravity benchmarks may potentially also be sensed at the network anchor, effectively reducing the observed relative difference and obscuring the gravity signal. Benchmark DYNG outside the caldera provides a more suitable anchor that is subject to less long-term vertical surface deformation. However, anchor VIKR is used here because it provides consistent and valuable information on the long-term gravity trend in a historical context. Future studies may consider using anchor DYNG for post-2022 surveys.

Drift corrections can be applied on a day-to-day basis (this study) as recommended by *Poland and de Zeeuw-van Dalfsen* [2019], or on a per-survey basis [*Giniaux*, 2019]. For example, an estimate of the linear drift rate over the full survey of 2016 provides an average rate of −470 µGal/d, while the average drift calculated over single campaign days ranges between −1230 to 558 µGal/d. While a gravimeter operating in a lab may experience continuous monotonic drift over days, changing environments appear to impose daily variations in drift rate [*Poland and de Zeeuw-van Dalfsen*, 2019]. Ideally, instrumental drift is estimated and corrected for on a per-circuit basis, but sometimes multi-day drift estimations cannot be avoided in the absence of anchor measurements (e.g., the 2017 campaign). For the data treated in this study, a daily drift correction is applied whenever possible, except for the 2017 campaign where a hybrid approach is employed.

Volcanic processes often produce minuscule gravimetric signatures and the choice of processing method has a demonstrable effect on the net microgravity changes. This effect may be amplified when a benchmark is only occupied once, as is the case for some measurements during the 2015 – 2017 campaigns. Single occupation measurements are particularly sensitive to variations in the assumed instrumental drift rate. This becomes especially critical when drift has to be estimated over multiple days due to e.g., the lack of consistent anchor measurements.

A campaign strategy using varying anchors between circuits can sometimes not be avoided, but it compounds the inherent uncertainty contained within the gravity measurements. It is recommended that each benchmark is measured at least twice at different times in a day, preferably in a double-loop or similar form. Multiple double measurements of benchmarks substantially help to accurately constrain the instrumental drift rate and vastly improve the confidence in the results. The adopted measurement strategy also depends on the type of instrument and for example, its susceptibility to drift, which may be lower for L&R instruments compared to modern Scintrex gravimeters. The precision and accuracy of the analysed microgravity data should be discussed. The precision of a measurement is naturally defined by its reported variance, but a second occupation with a different instrument, or on a different day, may produce a markedly different (albeit another precise) value. The measurements therefore show limited accuracy, despite the apparent high precision. This may happen for example after transport of the instrument, when insufficient time is available for the instruments to recover from tilting [*Reudink et al.*, 2014] before the measurement is made. Repeated measurements between two CG-5 instruments in the 2021 campaign show differences of 10 to 30 µGal, and up to 80 µGal on the worst occupations. Only for the data collected during the 2022 campaign, repeated measurements at benchmarks are usually consistent to within the reported measurement precision (sub-5 µGal for the CG-6 and sub-10 µGal for the CG-5). With only a single available measurement, the recovered value is generally considered to be representative. However, a second measurement of each benchmark (besides the added contribution of this measurement in determination of instrumental drift) provides critical insight into measurement repeatability. For the occupations with widely different reported values, measurements from two instruments were averaged – unless there existed a clear indication that one measurement was more reliable than the other (e.g., low versus high precision).

The exhaustive list of possible complications illustrates why the application of campaign gravimetry in volcano monitoring remains nontrivial. The limited number of benchmarks in the caldera, and high uncertainties of the data make it difficult to provide a quantitative analysis of the recovered microgravity results. However, the results that are presented in fig. 5.4 clearly indicate a long-term trend in the caldera center despite the observed yearly scatter, and also show that there is a definite coherency between measurements within regional groups. These results are discussed in the following sections.

5.4.2 Microgravity Signal During Subsidence (1988 - 2021)

The gravity record between 1988 – 2021 summarised in fig. 5.4 is likely partially influenced by the process driving the subsidence of the caldera center [*Pagli et al.*, 2006; *Sturkell et al.*, 2006; *de Zeeuw-van Dalfsen et al.*, 2013; *Giniaux*, 2019], acknowledging that e.g., hydrological changes may also have had a gravity effect. Generally over this period, a net microgravity decrease can be observed at the central benchmarks, suggesting a subsurface mass decrease. This observation is consistent with caldera subsidence as mass eviction is intuitively associated with deflating source volumes. Several source mechanisms have been suggested such as contraction of the magma chamber by cooling, and the removal of magma to lower levels [*de Zeeuw-van Dalfsen et al.*, 2013]. Indeed, seismic tomography from before 2013 [*Mitchell et al.*, 2013] reveals a shallow (< 3 km b.s.l.) high velocity anomaly below the caldera, indicating elevated densities and potentially, a contracting magma chamber, but may also represent a core of denser magma deposited in post-glacial time [*Brown et al.*, 1991; *Mitchell et al.*, 2013]. It is worth noting that the Bouguer survey conducted by *Brown et al.* [1991] is different from the analysis of temporal gravity changes explored here. The gravity decrease associated with subsidence may have ended in 2017 or continued until the uplift started in August 2021, but the precise ending cannot be determined confidently due to the lack of campaigns between 2017 – 2021.

The long-term decrease of the net microgravity was potentially interrupted in 2007 – 2008 and 2016 – 2017, when despite ongoing subsidence, an increase in net microgravity was observed in the eastern and center benchmarks. This observation was explained by a rising steam cap in the hydrothermal system, or magma inflow below the caldera [*Rymer et al.*, 2010; *de Zeeuw-van Dalfsen et al.*, 2013]. However, these minor deviations may be considered noise in the long-term decreasing trend or represent an unidentified effect of the same process that went undetected by alternative geodetic techniques.

5.4.3 Microgravity Change Leading up to and During Uplift

Increases in microgravity suggest that mass has been accumulating below the center of the caldera between 2016 – 2021 in the period leading up to the uplift. This almost negated the integrated effect of the previous two decades of observed gravity decrease (fig. 5.4). However, the 2021 campaign was completed while the uplift at Askja started an estimated 5 to 6 weeks earlier. The detected gravity increase of 100 µGal relative to benchmark VIKR – and potentially, a mass increase – may have played a role in causing the observed uplift. Unfortunately, the gravity increase cannot be accurately timed due to the lack of (reliable) microgravity campaigns between 2017 – 2021. This could mean that **i)** the mass increase could have occurred right before or during the start of the uplift in August, or **ii)** well in advance, but the magma has been filling up the available void space below the caldera first producing no detectable surface deformation, or **iii)** the caldera does not respond elastically to the magmatic source.

The gravity change between 2017 – 2021 is only detectable in the group of center benchmarks (fig. 5.4). This indicates that the source cannot be located at great depth with a large associated mass change. Such far-field source would have been sensed by multiple benchmarks in an broader spatial area that is not observed. However, the range of plausible source parameters for depth and mass that match the observations can be modeled (fig. 5.5). Parameters of the source likely surround the 100 µGal contour of OLAF and intersect with contours falling within the uncertainty limit of the measurements at MYV2 (estimated at 40 µGal), where no signal is detected. It is therefore likely that the source is shallower than 2.5 km below the surface, and does not exceed a mass change of roughly 1×10^{11} kg.

The microgravity campaign in 2022 was completed in an attempt to capture a hypothesized microgravity increase (i.e., mass inflow) associated with the observed uplift. Since August 2021, uplift appears to be centered west and below the center caldera benchmarks (fig. 5.6). However, the microgravity results show that there was no detectable increase in gravity between 2021 and 2022 after correcting for vertical deformation (fig. 5.4). Despite the apparent negligible change, high uncertainties in the 2021 campaign likely obscure any deep mass variations. This observation can be considered peculiar because Askja caldera continues to experience uplift after 2021 while there appears to be no detectable change in subsurface mass. It may thus be that subsurface mass emplaced during 2016 – 2021 is responsible for the progressive uplift observed during 2022. Perhaps the accumulation of mass occurred near the

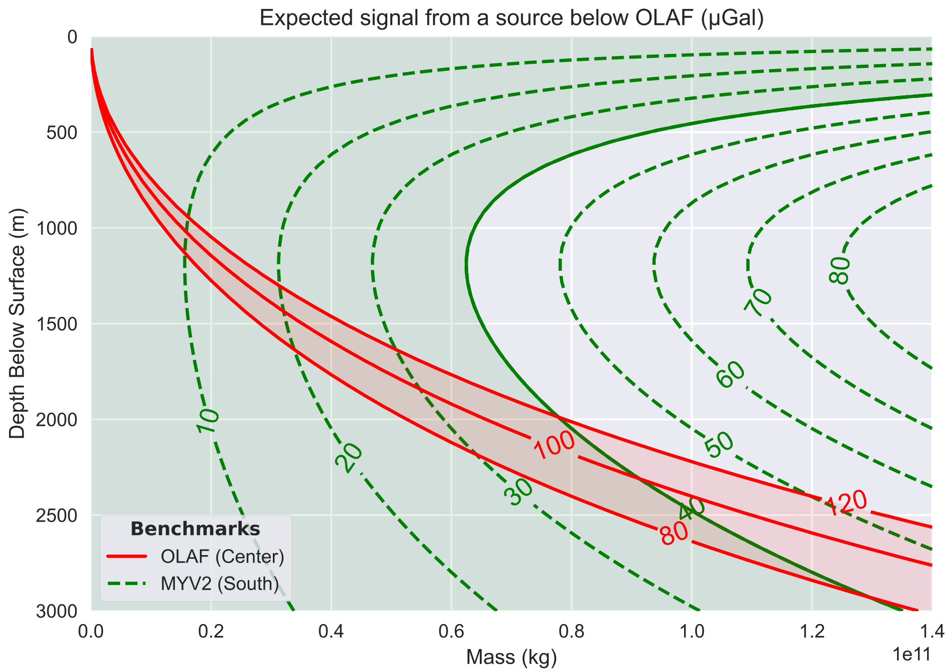

Figure 5.5: *Contours of the expected vertical microgravity change (μGal) for a point source at benchmark OLAF (red) and the closest non-center benchmark MYV2 (green) for a source directly below OLAF as a function of mass and depth (relative to VIKR). The solid green contour at 40 μGal represents an upper limit at which a signal would have been identified at MYV2, but was not. The intersection of the red and green areas represent the region of plausible source parameters.*

end of the period of subsidence, but did not induce any apparent response on the surface. Mass inflow with a muted surface response (i.e., without deformation) may be caused by a viscoelastic response of the crust to pressurisation at depth [*Zurek et al.*, 2012], accommodated by resident gas-rich magma [*Rivalta and Segall*, 2008], or through the filling of voids [*Johnson et al.*, 2010]. Such a process has been detected for example at Kīlauea, Hawaii where a net microgravity increase was observed before the onset of uplift [*Poland et al.*, 2019]. Drainage of a magma body below Askja caldera between 1988 – 2016 may have created voids and cracks that could have initially been filled during 2016 – 2021 before surface deformation was detected in August 2021. The uplift after August 2021 associated with a change in gravity following the free-air gradient suggests a different process involving density decreases in the subsurface.

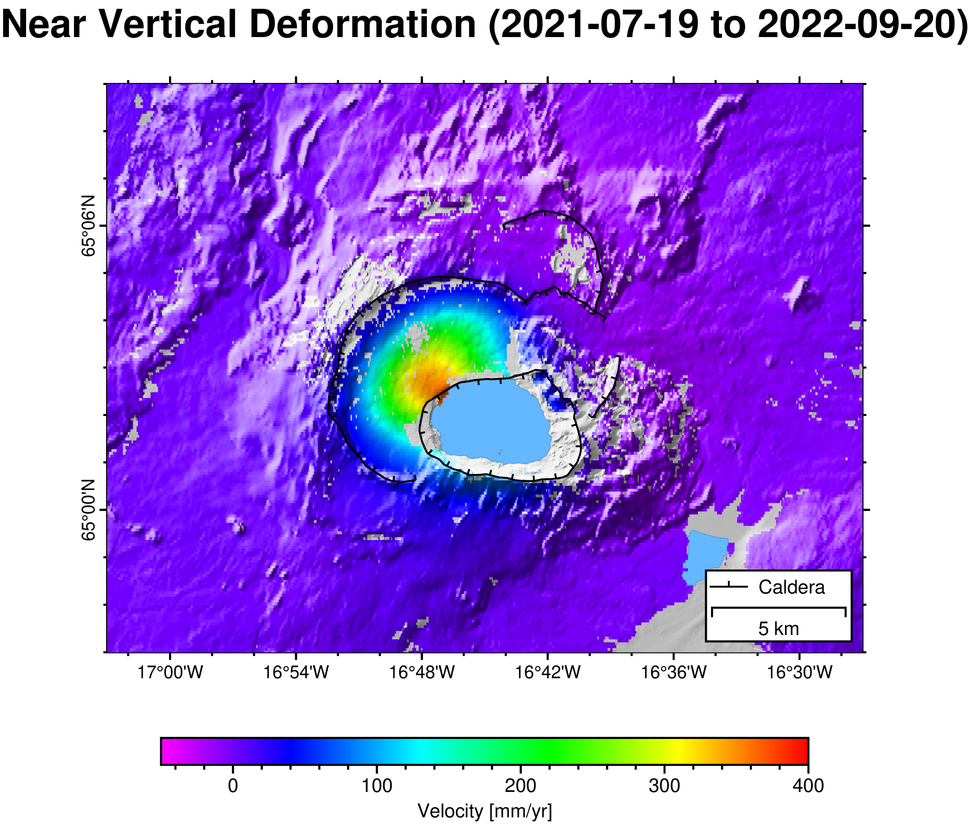

Figure 5.6: *InSAR deformation pattern from Askja between July 2021 and September 2022 showing the deformation rate in* mm/yr. *The concentric uplift pattern is clearly visible near the center of the main caldera.*

■ 5.4.4 Insight into the Driving Volcanic Processes

These microgravity results, when combined with the observed deformation patterns can be interpreted in terms of volcanic processes. One peculiar aspect of the recent uplift remains the limited amount of earthquakes below Askja caldera – features that are commonly associated with uplift (e.g., [*Sturkell et al.*, 2003]) but only if the stresses experienced during subsidence are exceeded [*Heimisson et al.*, 2015]. Weak and intermediate seismicity is not unusual for Askja [*Einarsson and Brandsdóttir*, 2021], but during a period of significant uplift, an increase in seismicity may be expected from bulging, although that does not directly infer magma movement [*Grapenthin et al.*, 2022]. Between 2016 and 2021, no significant increase in seismicity was detected near the storage reservoirs beneath the caldera [*Greenfield et al.*, 2020; *Winder*, 2022] – although seismicity associated with the ring faults surrounding the caldera increased during the period of uplift (T. Winder, personal communication). Overall, the absence of associated earthquakes may not be surprising considering that the co-eruptive and post-eruptive processes from the Holuhraun eruption in 2014 – 2015 [*Pedersen et al.*, 2017] may have already relieved excess stress in the subsurface. Alternatively, radiating heat from an intruding source may provide a ductile regime in the subsurface that is not characterised by brittle faulting.

In summary, the gravity data indicate that Askja appears to have experienced an inflow of mass somewhere during 2016 – 2021 and is now displaying uplift through a process that introduced no further mass below the caldera (i.e., a potential density decrease). A few potential hypotheses that fit these observations are discussed and may be considered:

Firstly, simple single-source geodetic models (M. Parks, personal communication) of the recent period of uplift between 2021 – 2022 indicate an estimated volume change of 0.013 to 0.018 km^3 at 1.3 to 2.9 km depth below the surface. With a nominal magma density between 2300 to 2700 $\mathrm{kg\,m^{-3}}$ and the assumption of a point source, this would produce gravity signatures within a range of positive 20 to 90 µGal directly above the source. The lower bound of this estimate remains within the uncertainty introduced by the 2021 campaign and such intrusion may thus remain hidden within the noise. Furthermore, the transfer of magma from a deep crustal volume to another shallower volume would cause little gravity change if the volumes are spatially adjacent.

Secondly, the microgravity changes observed during uplift fall around the free-air gradient, and all gravity decreases in the raw observations can more-or-less be attributed to changes in height. Effectively, this observation indicates that a volume change is experienced, without a significant increase in subsurface mass, and hence, a density decrease may be responsible for the uplift. Processes that induce volume changes without changes in mass may be magma vesiculation (bubble-forming) through e.g., the contact of a mafic intrusion with cooler rhylotic magma [*Eichelberger*, 1980], second boiling of intrusions over longer time scales [*Wech et al.*, 2020], a viscoelastic response of the crust to pressurization at depth [*Zurek et al.*, 2012], or possibly the formation of voids [*Gottsmann and Rymer*, 2002; *Van Camp et al.*, 2017]. Considering the influx of mass between 2016 – 2021, the formation of voids seems unlikely, and bubbles forming from magma migrating to shallower levels may be a more suitable mechanism to explain the observations.

A third process that can produce uplift without concurrent subsurface mass accumulation would be a change in the hydrothermal system. Heat coming from the newly intruded material between 2016 and 2021 could have an effect on the extensive hydrothermal system at Askja. It is not uncommon that a process like this is expressed as significant temporal geodetic changes at volcanoes [*Saibi et al.*, 2010]. The phase transition from liquid water to steam is commonly detected at around 2 to 3 km below the surface [*Greenfield et al.*, 2016; *Halldórsdóttir et al.*, 2010], which may be consistent with the recovered source from surface deformation measurements. Furthermore, the center of deformation is located on the western edge of lake Öskjuvatn, around which hydrothermal activity is commonly observed [*Ranta et al.*, 2023]. Future observations of increased surface heat flow or a similar increase in the temperature of the lake may support this hypothesis. Furthermore, heat from the intrusion may allow the host rock to respond in a ductile instead of brittle fashion, explaining the lack of apparent increased seismicity.

A final possibility to be considered is the replacement of denser basaltic magma ($\rho = 2800\,\mathrm{kg\,m^{-3}}$) with its rhyolitic ($\rho = 2400\,\mathrm{kg\,m^{-3}}$) counterpart. This interpretation would have implications of the hazard assessment at Askja as rhyolitic magmas are often associated with more destructive eruptions than basaltic magmas. Alternatively, it may be caused by a convecting basaltic body below the caldera where gas-rich basaltic magma is replacing denser degassed magma at shallow depth. In the case that such process is driving the uplift, surface deformation data from 2023 – 2024 will continue to show uplift with gravity changes following the free-air gradient.

The driving mechanism for the uplift remains enigmatic and further microgravity campaigns in 2023 and 2024 will help shed light on the cause of the activity. It is recommended such microgravity campaigns are completed.

5.5 Conclusion

Since 1988 microgravity measurements at Askja show that the extended period of subsidence coincides with an observed decrease of around 100 to 150 µGal restricted to the center of the caldera. Microgravity differences between 2016 – 2021 indicate a significant increase (100 to 120 µGal) in gravity associated with mass accumulation in the center of the caldera occurred either during a period of subsidence before 2017, or leading up to or during the period in inflation after August 2021. Due to the lack of microgravity campaigns between 2017 – 2021, the mass increase cannot be more accurately timed. The recent period of uplift appears characterised by an insignificant change in microgravity, although a signal may still remain hidden due to the elevated uncertainty (45 µGal). Although the unambiguous interpretation of microgravity in terms of volcanic processes remains challenging, these results indicate that the surface deformation being detected in 2021 – 2022 may be a consequence of mass emplaced somewhere between 2016 and 2021. The previously emplaced mass may play a role in causing the observed uplift without the accumulation of additional mass, and likely indicates a process that involves subsurface density decreases such as magma vesiculation, a change in the hydrothermal system, the replacement of denser basaltic magma with less dense silica-rich magma, but may also represent viscoelastic or poroelastic relaxation of the subsurface.

5.6 Acknowledgements

Erik Sturkell, Sveinbjörn Steinþórsson, and Catherine O'Hara are thanked for facilitating and joining the geodetic fieldwork at Askja. We thank the department of Geoscience & Remote Sensing of TU Delft and the University of Leeds for providing the Scintrex CG-5 and CG-6 instruments used during the 2021 and 2022 campaigns. We are grateful to one anonymous reviewer and Glyn Williams-Jones who diligently helped to improve the manuscript. Additionally, Henrietta Rákóczi is thanked for her gratuitous review of our manuscript. Figures were made with

Matplotlib [*Hunter*, 2007], version 3.2.1 [*Caswell et al.*, 2020]. This research was supported by the EU Horizon 2020 EUROVOLC project funded by the European Commission (grant number 731070), the H2020 DEEPVOLC project funded by European Research Council (grant number 866085), and the FETOPEN-2016/2017 call (grant number 801221). R.G. acknowledges partial support through NSF grant EAR-1464546. COMET is the UK Natural Environment Research Council's Centre for the Observation and Modelling of Earthquakes, Volcanoes and Tectonics, a partnership between UK Universities and the British Geological Survey.

5.7 Data Statement

Deformation data are courtesy of the University of Iceland and available via EarthScope (formerly UNAVCO). Microgravity campaign data from the 2021 and 2022 campaigns are available from 4TU [*Koymans et al.*, 2023]. Microgravity data between 2015 – 2017 were downloaded from the Leeds Data Research Repository [*Giniaux*, 2021]. Hillshaded map data were downloaded from the *National Land Survey of Iceland* [2020]. Geographical data include geological features [*Johannesson and Sæmundsson*, 2009], volcanic zones [*Jakobsson et al.*, 2008], and glaciers [*Náttúrufræðistofnun Íslands – Icelandic Institute of Natural History*, 2018].

6

Synthesis

In this final chapter the material presented in the dissertation is reviewed, and it also serves to provide a high-level synthesis of the preceding four chapters. A detailed discussion and conclusions of the individual chapters are naturally included with the respective chapters. The purpose of the synthesis is to provide an encompassing overview of conclusions, shortcomings, and recommendations, including an outlook for potential future work.

6.1 Conclusions and Recommendations

Volcanic eruptions are inherently hazardous to society. Continuous monitoring of active volcanoes therefore remains an important task in the protection of civilians in order to avoid the loss of lives and livelihood. This dissertation contributes to the incremental advancement of volcano monitoring techniques through **i**) the performance assessment of instruments commonly employed in volcano monitoring networks (section 6.1.1) and **ii**) the use of campaign microgravity in long-term volcano monitoring. A recurring theme of this dissertation concerns the topic of data quality, of which the evaluation is crucial if geophysical data is to be used reliably for volcano monitoring. Emerging technologies in volcano monitoring using MEMS technology, and the use of campaign microgravity are discussed (section 6.1.2). This section includes many recommendations on how microgravity campaigns can be completed and treated to improve the derived reliability of results (section 6.1.3). In section 6.1.4 the integration of continuous gravity data with existing FAIR repositories is discussed – a prerequisite for microgravity data to be used effectively in an operational volcano monitoring infrastructure.

■ 6.1.1 Performance Assessment of Geophysical Instruments in Volcano Monitoring Infrastructures

Assessment Using Power Spectral Density Estimates

Multi-hazard assessment and the monitoring of active volcanoes requires a cross-disciplinary approach, integrating data and knowledge from various geoscientific fields. For example, seismometers generally provide a measure of the amplitude of volcanic tremor and the information on the amount, magnitude, and location of different types of volcanic earthquakes. Geodetic observations such as tilt, leveling, GNSS and InSAR data may contribute additional insight into surface deformation and subsurface pressure and volume changes. These observations can be complemented by constraints on subsurface mass and density changes from microgravity observations. Other techniques, such as visual observations, measurements of gas emission, or acoustic detection provide alternative, valuable observations. Because of the varied nature of instruments employed in volcano monitoring networks, chapter 2 focuses on the quality assessment of eight types of geophysical instruments using a standardized, automated analysis of power spectral density (PSD) estimates. The proposed technique can be effectively applied to many types of geophysical instruments that produce time series and is demonstrated on data from accelerometers, geophones, broadband seismometers, tiltmeters, microbarometers, hydrophones, superconducting relative gravimeters, and GNSS receivers. The described approach is employed in an operational system by the Royal Netherlands Meteorological Institute (KNMI) to detect the degraded performance of accelerometers, geophones, broadband seismometers, and microbarometers that are deployed in the field. The presented work builds on existing methodologies (e.g., *McNamara and Boaz* [2006a]) and generates PSD estimates on a daily basis for freshly archived data. These data are verified against novel theoretically derived and data-driven statistical constraints, such as long-term historical observations at an installation site. Through this approach, instrumental problems have been detected shortly after they commenced which may otherwise have gone unnoticed for an extended period of time. It is recommended that such an approach – or one comparable – is implemented by other monitoring institutes and optionally extended with additional constraints that can be imposed on the PSD estimates.

Passive Assessment Using Variations in the Electrical Network Frequency

Chapter 3 introduces a complimentary assessment technique for instrument performance that was adopted from audiovisual forensic analysis (e.g., *Cooper* [2008]). This chapter demonstrates the use of recorded minor fluctuations of the electrical network frequency (ENF) of 50 Hz (in Europe) in geophysical readings as a nationwide calibration signal. It is shown that using this signal*, timing corrections can be applied to instruments with a resolution of roughly one second. This technique may be valuable to correct geophysical instruments that rely on timing information

*Attentive readers may have identified the subtle (overlooking this rather conspicuous footnote) change from *noise* to *signal* since the introduction.

through e.g., the less accurate network time protocol (NTP) compared to time information derived from GNSS. In a completely separate approach, it is shown that the electrical grid frequency can also be used as a calibration signal from a stable unchanging source, and can potentially be leveraged to detect orientation errors of seismometers.

6.1.2 Emerging Technologies in Terrain Gravimetry and Volcano Monitoring

Progress towards real-time monitoring of active volcanoes using microgravity remains challenging because it is difficult to continuously obtain gravity data and subsequently identify precursor signals before disruptive volcanic events transpire. At present, continuous microgravity observations have only rarely been applied in forecasting (e.g., after the detected microgravity increase in 2019 at Kīlauea [*Poland et al.*, 2019]) to identify potential precursors leading up to eruptive events. In other words, these observations were not completed in real-time, but after the disruptive volcanic event had already occurred (e.g., lava fountaining [*Carbone et al.*, 2013] or caldera collapse [*Poland et al.*, 2019]), providing a clear direction where to search for such potential precursor signals. One key achievement would be to further integrate continuous gravity observations in real-time monitoring infrastructures and continuously search for microgravity signals that could indicate imminent volcanic activity, as is done for e.g., Mt. Etna [*Carbone et al.*, 2019].

The lack of real-time integration of microgravity data in operational systems is a shortcoming that is mainly a consequence of the limited availability of gravimeters. This constraint is generally imposed by high instrumental cost. Gravimeters need to be extremely sensitive and stable, so much that they are extremely difficult to design, let alone produce commercially at a reasonable price. The most accurate and precise instruments in particular (i.e., relative superconducting gravimeters and absolute quantum gravimeters) are usually very costly and require supporting infrastructure, such as a mains power supply and shelter*. With a limited number of available instruments, the consequence is a direct trade-off in resolution between measurements in space and time. In order to effectively leverage microgravity measurements in real-time volcano monitoring infrastructures, the deployment of multiple continuous instruments is thus a prerequisite. Emerging technologies and ongoing advances in terrain gravimetry include the development and adoption of cost-effective spring-based MEMS (Microelectromechanical Systems) gravimeters (e.g., *Middlemiss et al.* [2016]). Such instruments could provide a high number of spatially distributed continuous microgravity observations in a cost-effective way. Naturally, the implicit assumption in such strategy would be that the added value in spatio-temporal resolution directly outweighs the lower expected performance of the instruments compared to that of conventional, more expensive instruments.

*It may be unsurprising that these requirements are often nontrivial to find near the summit of active volcanoes.

The NEWTON-g Gravity Imager

The deployment of such MEMS gravimeters was explored in the FET-2016 project NEWTON-g, that aimed to install a so-called gravity imager on Mt. Etna. The imager was imagined using 20 – 30 MEMS gravimeters that would represent pixels of the imager, including a single commercial absolute quantum gravimeter (AQG) as reference point. The dissertation was initially conceptualized in the framework of that project to leverage the increased spatio-temporal resolution of microgravity observations in volcano monitoring. Data from the completed gravity imager could lead to advance the detection of continuous microgravity precursor signals in volcano monitoring, eventually evaluating the implications for hazard and risk assessment. Due to unforeseen circumstances along the way that can be mainly attributed to the COVID-19 pandemic and imposed restrictions by national governments, only a limited number of MEMS gravimeters could be installed over the course of the project. Furthermore, the four MEMS gravimeters that are presently (as of March 2023) operational in the field appear susceptible to fluctuations in ambient pressure and temperature – potentially caused by minuscule punctures in the seal surrounding the sensor. These issues have proven challenging to eliminate both through applied hardware upgrades, as well as passively through signal processing. The environmental fluctuations cause the measurements of the instruments to drift unpredictably over the long-term, making it challenging to distinguish between volcano related and instrumental related signals. However, the pursuit is still ongoing and many improvements have been made to the MEMS gravimeters based on the valuable data collected from Mt. Etna over the course of the NEWTON-g project [*Anastasiou et al.*, 2022]. More importantly, despite the present shortcomings, short-lived events with large amplitude gravity changes (e.g., lava fountainig [*Carbone et al.*, 2015]) may still in fact be detectable, because the MEMS gravimeters in the field are capable of detecting gravity signals stemming from the solid Earth tides. Other activities in the project that were accomplished include the deployment of an absolute quantum gravimeter (AQG) near the summit of Mt. Etna at 2800 m altitude [*Antoni-Micollier et al.*, 2022]. As a consequence of the lack of available MEMS gravimeters and sufficient microgravity data to work with effectively, the scope of the dissertation was diverted and mainly written to concern the use of campaign microgravity in long-term volcano monitoring. Campaign microgravity remains extremely relevant to terrain gravimetry as well as the NEWTON-g project, since the gravity imager would rely on campaign measurements to anchor the MEMS gravimeters to the reference point provided by the AQG. In volcano monitoring, it is evident that a hybrid microgravity approach [*Hinderer et al.*, 2016] is extremely valuable as it combines the strengths of absolute & relative, and campaign & continuous recordings.

The Role of Campaign Gravimetry in Volcano Monitoring

Campaign microgravity measurements exist as a historically proven method for the detection of gravity signals associated with long-term volcano uplift or subsidence. Campaigns are completed using portable instruments (e.g., Scintrex CG-5, CG-6, and Burris gravimeters) that are convenient to carry in the field and usually provide a single average measurement per site on an episodic (e.g., yearly) basis. However, these instruments naturally suffer from problems like aliasing from short-period transient gravity signals (e.g., fluctuations in groundwater level or snow [*Poland and de Zeeuw-van Dalfsen*, 2019; *Carbone et al.*, 2019]) that were present during the campaign, and have a high sensitivity to the applied campaign strategy (e.g., single benchmark occupations per campaign) that should be reduced as much as possible. The effect of environmentally induced gravity changes are very challenging to estimate accurately and long-term trends should be considered, while outliers from a single campaign may not be considered representative. Campaign microgravity measurements also suffer from a legacy problem on how the network of benchmarks was initially designed, potentially up to decades ago. Often it is decidedly better to continue measurements of an existing network of benchmarks than it is to (partially) re-design and lose sometimes over a decade-long consistent data series. When designing a gravity network, benchmarks should preferably be co-located with continuous or campaign GNSS benchmarks in order not to be forced to rely on less accurate (or unavailable) InSAR data. Furthermore, benchmarks should be designed with clear markers for adjustable tripods that support the gravimeters.

During the writing of this dissertation it became apparent there is an evident need for a discussion on a fundamental level on how campaign microgravity data should be collected and treated. A lot of effort can be put into correcting transient minor microgravity effects (e.g., atmospheric pressure, water and snow level variations) in order to reduce short-period aliasing effects. However, this effort can be considered mostly in vain if a more significant systematic unknown is introduced by an ineffective campaign strategy or analysis. The focus of chapters 4 and 5 is on the use of microgravity campaign data in the detection of long-term volcanic processes at Kīlauea and Askja respectively. Beside the interpretation of gravity results in relation to the observed surface deformation, there is a strong focus on recommendations for improvements that can be made to the collection and treatment of data, that can immediately benefit the reliability of microgravity campaign results. Microgravity data can be considered reliable if campaigns are completed in an appropriate manner and data are treated properly. It is critical that benchmarks are measured twice in a circuit, using two instruments, and preferably repeated in two circuits. The reliability of microgravity measurements is often discussed and the analysis are often limited to being qualitative due to the inherent high variance in data. (Continuous) gradient measurements may also help to improve results – specifically when considering the effect that deformation has on the gravity results. The theoretical free-air gradient correction that is applied may not be representative in all cases. Such observation would also assist with joint inversions of surface deformation and microgravity data, which can be applied if we assume they are caused by the same source. For this reason, the following section presents practical recommendations that can be implemented to improve the effectiveness of microgravity campaigns.

6.1.3 Recommendations for Successful Gravimetry Campaigns

In the treatment of two extensive campaign microgravity records (chapters 4 and 5), a number of strong points and pitfalls were identified in the applied campaign strategies and conventional analysis method. Hence, the following set of recommendations were designed so that they can be easily adopted (or continued) during the collection and treatment of campaign microgravity data to directly improve the reliability of the results. Consider the following *fourteen* commandments of campaign microgravity:

[1] Microgravity campaigns should be completed at regular intervals and not just in response to disruptive events because gravity changes leading up to the event may have passed undetected.

[2] Instrumental drift should be estimated and reduced on a per-circuit basis and not over a full campaign. This should eliminate any (subtle) variations in instrumental drift that are observed during different campaign days (e.g., *Poland and de Zeeuw-van Dalfsen* [2019]).

[3] Single looped circuits are inherently insufficient because the accuracy of single microgravity measurements can not be independently verified. Besides, an incorrectly applied tidal correction or instrumental tare will pass undetected in a single loop.

[4] Double loops, or a modified version thereof are recommended, providing a minimum of two occupations of each benchmark per circuit that will provide insight into measurement repeatability and contribute to accurately constraining instrumental drift rates and data tares.

[5] Two occupations of the same benchmarks should be made with the maximum possible time duration between them in order to assist with the estimation of instrumental drift rates.

[6] Instrumental drift should be determined on basis of individual measurements during an occupation (and not an average) because even sequential measurements include information on the instrumental drift rate.

[7] In order to further verify the accuracy of microgravity measurements, the campaign should preferably be completed with two independent instruments. Without two instruments available, it is recommended that each benchmark is occupied on different days in separate double loops.

[8] Gravity corrections for ocean loading and polar motion [*Wenzel*, 1996] often do not contribute significant improvements over the comprehensive tidal formulations by [*Longman*, 1959]. These corrections are only relevant for high-precision (sub-µGal) gravity measurements.

[9] Microgravity differences with the network anchor should be recovered during analysis in a single inversion, simultaneously with (linear) instrumental drift parameters including an optional tare [*Reilly*, 1970; *Hwang et al.*, 2002; *Hector and Hinderer*, 2016; *Koymans*, 2022a]. See appendix A of this dissertation for the implementation details.

[10] Microgravity campaign data can be inverted for data from two distinctive instruments simultaneously.

[11] It is therefore always preferable to install two co-located microgravity benchmarks that can be measured by two instruments at the same time.

[12] Benchmarks should be occupied for at least ten minutes (e.g., in 2 × 5 minute intervals), and possibly extended when the measurement appears unstable. This is critical for instruments like the Scintrex CG-5 and CG-6 that are susceptible to tilt and must recover after transport. These instruments converge towards a stable value eventually but may take up to 20 min [*Reudink et al.*, 2014].

[13] Measurements that are not made in the same loop as the network anchor should be avoided because the measurement uncertainties are compounded. If one poor measurement is made at a substitute anchor this can throw off all data when they are expressed relative to the conventional network anchor (e.g., see chapters 4 and 5). If this cannot be avoided, substitute anchors should definitely not be selected in places with transient gravity changes (e.g., close to an active lava lake).

[14] Microgravity campaign data should be published in public data repositories in the standard instrumental output formats to facilitate software interoperability.

Naturally, logistics and weather often prevent campaigns to be completed in an optimal fashion. However, as demonstrated by data presented in this dissertation, particularly by the campaign completed at Askja in 2022, microgravity campaigns can indeed produce repeatable and reliable results on the sub-5 µGal level. Potentially, further improvements that can be made are to measure close to GNSS benchmarks to get accurate deformation constraints, and to measure changes in free-air gradients at the benchmarks through time – but these recommendations are often admittedly not feasible to complete or trivial to implement. Hence they are not included as part of the presented list of recommendations.

6.1.4 Integration of NEWTON-g Microgravity Data with the European Integrated Data Archive

The integration of microgravity data with existing data infrastructures returns to the desire that is expressed at the start of this synthesis. Volcano monitoring requires a multi-disciplinary approach, and that expresses a need for shared data infrastructure. In addition to the collection and treatment of microgravity data, the gravity community should consider dedicating a proportional amount of time to the adoption of standards offered by existing modern data infrastructures. Data repositories that are specifically designed for gravity data exist (e.g., the International Gravimetric Bureau, BGI), but have evolved less over the past decades compared to available alternatives. For example, ORFEUS – or more generally – The International Federation of Digital Seismograph Networks (FDSN)) represents a community that is extremely well organised. ORFEUS has been effective at promoting open and FAIR data sharing for decades in initiatives such as the European Integrated Data Archive (EIDA). ORFEUS originated from the seismological community and has more than a decade lead in the development of FAIR compliant data management and dissemination services. It provides modern services that make data findable through metadata, and subsequently easily and freely accessible through web services. These infrastructures also leverage protocols that support real-time data acquisition (e.g., Seedlink), which is critical for the downstream use of microgravity data in volcano hazard assessment.

The NEWTON-g consortium collaborated with ORFEUS to integrate data from the NEWTON-g gravity imager with one of its existing data archives (EIDA). The data collected by the AQG between August to December 2020 are publicly available, with the rest of the data remaining under an embargo until 2024. Two software tools were created to facilitate the integration with EIDA namely AQG2MSEED [*Koymans*, 2022b] and WEEG2MSEED [*Koymans*, 2022c] that convert data output files from the instruments to the mSEED standard format. Presently, the conversion is done episodically, and data is not being streamed in real-time into the archive. This shortcoming would require a joint effort including instrument manufacturers to implement real-time streaming protocols and should be considered in the future.

6.2 Outlook

This section completes the synthesis and offers some perspective for future work. When designing or upgrading a microgravity network, more consideration should be given to its design so that many decades later it can still be used. Additionally, for campaign gravimetry, consistency is key, and for the benefit of future generations of scientists it may be worth investing the extra time right now. The continued development of cost-effective MEMS gravimeters is valuable, because increasing the spatio-temporal resolution remains one of the most sought-after improvement in microgravity analysis. In recent years, large leaps have been made but this will remain an ongoing challenge for the coming years. While campaign gravimetry is evidently a useful tool for the detection of long-term and large wavelength processes causing gravity change (fig. 1.1), the added benefit of continuous observations should not be understated. These observations would be valuable to detect transient gravity signals associated with e.g., lava fountaining [*Carbone et al.*, 2013]. A hybrid microgravity approach is recommended – leveraging the strengths of both absolute and relative, and campaign and also continuous microgravimetry.

6.2.1 Joint Inversion of Gravity and Surface Deformation

Another challenge that is actively being worked on in the literature concerns the identification of deformation-induced gravity changes. In microgravity analysis it is conventional that an a-priori correction to the microgravity results is made for vertical deformation that follows the free-air gradient before source inversions are completed. However, it is clearly better that these parameters are considered together, which is not done in this dissertation and represents a clear shortcoming. Recent studies and models are being developed for the joint inversions of deformation induced height and gravity changes since both geodetic measurements contain information about the source of change, assuming this source is the same. Different sources (e.g., a sill, prolate, or a Mogi source) may produce similar surface displacements, but could be distinguished by integrating the information from microgravity data [*Nikkhoo and Rivalta*, 2022]. It should be noted that, as is also demonstrated by chapters 4 and 5 in this thesis, that there may be an incongruity between surface deformation and subsurface mass accumulation or withdrawal (e.g., void creation and filling). Hence, an elastic model may not necessarily be directly applicable in all volcanic settings. Furthermore, a significant number of reliable microgravity readings with a wide spatial distribution is required, and that is often difficult to obtain. Such joint inversion would however be excellent where **i)** sufficient deformation and microgravity data are available, and **ii)** the volcano responds to an intrusion in an elastic manner. Joint inversions of gravity and deformation would greatly help reduce the ambiguity that often exist in the determination of the source. Campaign microgravity measurements may lend themselves to being very suitable for this because they provide a high spatial resolution over the long-term. The heart of the problem is that a higher spatial coverage is required, and more importantly, for any model whatsoever, the reliability of input data constrains the reliability of the output. In practise, microgravity data with low resolution in space and low data quality lend themselves to be better interpreted qualitatively instead of quantitatively. Reliable quantitative estimates are often preferable but are sometimes hard to achieve. Regardless, even qualitative long-term trends such as displayed by the benchmarks in the center caldera of Askja in fig. 5.4 can hardly be disregarded, despite some often unquantified residual in the data.

6.2.2 Applications of ENF Analysis

The technique may potentially assist in the passive determination of unconstrained sensor orientations of geophones inside seismic boreholes. The technique appears promising, and further research on this aspect may be considered, particularly on how the ENF signal is coupled to different types of instruments. Other discovered potential applications may reside within variations in the total power that is coupled to the instruments in a narrow band around the ENF that also appears to vary synchronously nationwide. Furthermore, other advantages of the technique that are naturally borrowed from forensic analysis is that tampering of geophysical data can potentially be detected, which may serve a purpose in the verification of e.g., the Comprehensive Nuclear-Test-Ban Treaty [*Coyne et al.*, 2012]. An avenue for future work is suggested to be the identification of the coupling mechanism of the ENF to various geophysical instruments through experimental means.

While complex analyses can be made with geodetic measurements including microgravity observations, at the most fundamental level it starts with guaranteeing the reliability of microgravity results and that they are of the highest possible and achievable quality – which this dissertation hopefully contributes to with its recommendations and conclusions (section 6.1.3).

A

Derivation of the Gravity Adjustment Methodology

A.1 Introduction

This appendix provides an in-depth description of the implementation of the weighted least square (WLS) gravity adjustment routine. The method is similar to the datum free constraint [*Hwang et al.*, 2002] and applied to short double looped circuits. An additional optional degree of freedom can be added to the inversion in order to automatically fix suspected data tares. The following method is implemented as a web application [*Koymans*, 2022a].

A.1.1 Gravity Adjustment

The undesirable effect of instrumental drift must be considered before effective gravity differences between the benchmarks and the anchor can be reliably recovered. A linear function usually suffices to capture the largest component of thermal and mechanical drift that occurs within spring gravimeters. Clearly, this trend can be recovered through a linear regression using multiple observations at a single benchmark made at different times during the circuit, and is the main reason why circuits should be closed. A regression can be represented as a linear system of equations and is conventionally expressed as a standard inverse problem [*Menke*, 2018]:

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \boldsymbol{\epsilon} \tag{A.1.1}$$

Where $\mathbf{G}$ represents the design matrix that relates an unknown model vector $\mathbf{m}$ to an observation vector $\mathbf{d}$, with $\boldsymbol{\epsilon}$ being the residual vector. A solution for the unknown model vector can be estimated by minimizing the sum of weighted residuals squared between the model and the observations:

$$E = \boldsymbol{\epsilon}^{\mathrm{T}}\mathbf{W}\boldsymbol{\epsilon} = (\mathbf{d} - \mathbf{G}\mathbf{m})^{\mathrm{T}}\mathbf{W}(\mathbf{d} - \mathbf{G}\mathbf{m}) \tag{A.1.2}$$

with knowledge on the inversely weighted variance $\mathbf{W}$ of the observation vector $\mathbf{d}$ on the diagonal. Minimising the objective function E with respect to the model parameters evaluates to the classical normal equation. This particular solution of eq. (A.1.2) for the estimated best-fit model parameters $\tilde{\mathbf{m}}$ is given by:

$$\tilde{\mathbf{m}} = [\mathbf{G}^{\mathrm{T}}\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{W}\mathbf{d} \tag{A.1.3}$$

To illustrate, in order to estimate the instrumental drift during a circuit, given a number of n independent gravity observations g_{n} in µGal from a single benchmark, taken at times t_{n} (s) since the start of the circuit, the linear regression can be formulated following eq. (A.1.1) in matrix form:

$$\begin{bmatrix} g_0 \\ \vdots \\ g_{\mathrm{n}} \end{bmatrix} = \begin{bmatrix} t_0 & 1 \\ \vdots & \vdots \\ t_{\mathrm{n}} & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \boldsymbol{\epsilon} \tag{A.1.4}$$

Where the model vector $\mathbf{m}$ includes the slope (α) and intercept (β) of the linear regression. With sufficient independent measurements this problem becomes easily overdetermined with more equations (n) than unknowns (2). For the system in eq. (A.1.4), the model parameters (α, β) recovered from eq. (A.1.3) represent the parameters of straight line ($g = \alpha t + \beta$) that has the smallest squared residuals from the combined gravity observations. The slope of this line inherently represents the estimated drift rate during the campaign day in µGal per unit of time.

Evidently, eq. (A.1.4) is only applicable to gravity observations made at a single location when the value of gravity remains constant outside of the influence of instrumental drift. Without an absolute reference point, gravity observations are expressed relative to the network anchor. In relative campaign gravimetry, stepwise observations of gravity are measured at benchmarks that vary by a constant value $\Delta \mathrm{g}_{\mathrm{bench}}$, namely the effective difference in gravity between each benchmark and anchor. This constant component per benchmark can be introduced as an additional linear term in the basic system presented in eq. (A.1.4). The unknown model vector $\mathbf{m}$ then becomes a concatenation of the (linear) instrumental drift parameters (α, β) and gravity differences vector $\mathbf{m}_{\mathrm{g}}$:

$$\mathbf{m} = \begin{bmatrix} \alpha \\ \beta \\ \mathbf{m}_{\mathrm{g}} \end{bmatrix} \tag{A.1.5}$$

Where $\mathbf{m}_{\mathrm{g}}$ is a column vector that contains the set of unknown gravity differences between each benchmark and the anchor. Likewise, the design matrix $\mathbf{G}$ needs to be modified to include one additional linear term introduced per benchmark. The structure of each added binary column in $\mathbf{G}$ ensures that the constant gravity difference with the anchor is added to the gravity observations at the respective benchmark. The shape of $\mathbf{G}$ will be $\mathrm{N} \times \mathrm{M}$ where N is the number of total observations, and M is the number of benchmarks plus the chosen number of regression parameters. For an example circuit using a double closed loop approach measuring a single benchmark in addition to the anchor, the problem can be expressed as:

$$\begin{bmatrix} g_{\text{anchor},t_0} \\ g_{\text{bench},t_1} \\ g_{\text{anchor},t_2} \\ g_{\text{bench},t_3} \\ g_{\text{anchor},t_4} \end{bmatrix} = \begin{bmatrix} t_0 & 1 & 0 \\ t_1 & 1 & 1 \\ t_2 & 1 & 0 \\ t_3 & 1 & 1 \\ t_4 & 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \Delta \mathrm{g}_{\text{bench}} \end{bmatrix} + \boldsymbol{\epsilon} \tag{A.1.6}$$

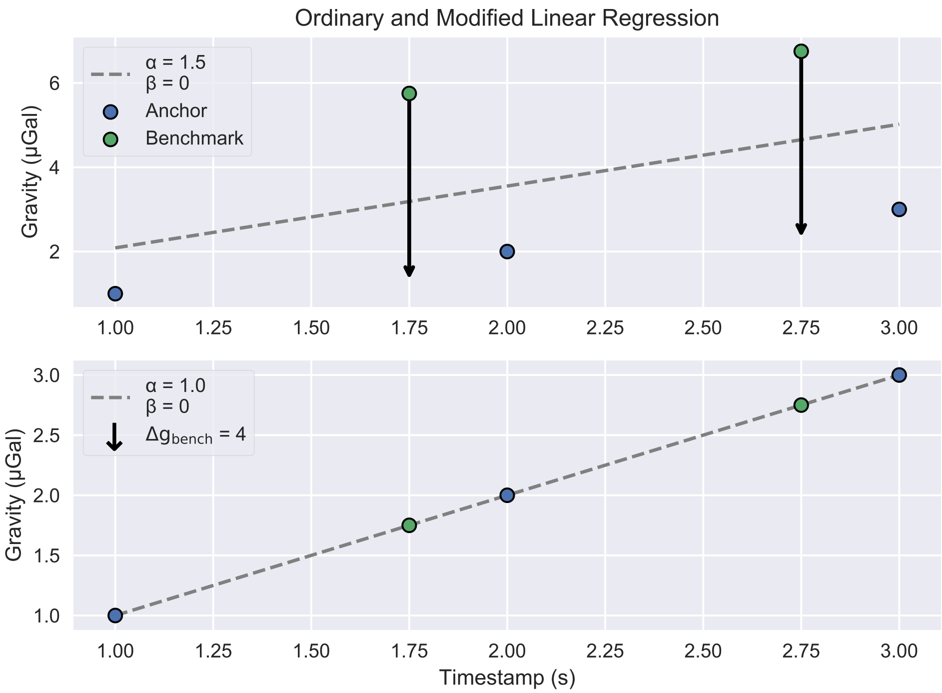

Figure A.1: *Top) basic linear regression following eq. (A.1.4) on all the observations naturally fails to capture the drift rate accurately because of the difference (black arrows) in microgravity between the benchmark and anchor. Bottom) With the addition of $\Delta \mathrm{g}_{\text{bench}}$ to the system in eq. (A.1.6), the data become aligned on a single best-fit line after elimination of the recovered gravity difference.*

The example system defined in eq. (A.1.6) is visually represented in fig. A.1, with synthetic observations from a circuit in a double closed loop form, consisting of a base station and single benchmark with a virtual gravity difference set at

4 µGal. A constant drift rate of 1 µGal/s is imposed on the observations. The figure shows that the gravity value at the anchor is observed at three instances (t_0, t_2, t_4), alternated with two gravity observations at the benchmark (t_1, t_3). The unknown gravity difference with the anchor Δg_{bench} is added to all observations made at the benchmark. Effectively, during the inversion defined in eq. (A.1.3), the collective group of data from the benchmark is given an additional degree of freedom to shift vertically and align itself on the best fit line with the observations at the anchor, implicitly accounting for instrumental drift. The best estimate for the model parameters can thus be recovered by applying eq. (A.1.3) to the system presented in eq. (A.1.6). This system can be extended to an arbitrary number of observations and benchmarks, as long as sufficient observations are present for the system to remain overdetermined. Data tares that were observed can be restored through an additional degree of freedom in $\mathbf{m}$, adding a constant offset to the collection of observations taken before or after the tare. The moment of the tare would need to be configured by the operator, but its offset would be automatically corrected during the inversion. An additional improvement could be made by integration data from two gravimeters (e.g., CG-5 and CG-6) by extending the design matrix in eq. (A.1.5) with independent drift parameters for each instrument (α_{CG-5}, β_{CG-5}, α_{CG-6}, β_{CG-6}).

A.1.2 Gravity Adjustment Parameter Uncertainty

The observation variance on the diagonal used in the weight matrix $\mathbf{W}$ implies uncorrelated observational errors and represents the standard error of the gravity measurements, hence a relative measure of confidence in each observation. It is calculated from the measurement variance σ_x^2 of each observation divided by the number of samples n used for the measurement:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} \tag{A.1.7}$$

For the CG-5, the number of samples is calculated from the measurement duration in seconds multiplied by the instrument sample rate of 6 Hz, minus the number of rejected samples during the observation. This uncertainty estimate does not entirely capture the residual between the data and the model $\boldsymbol{\epsilon}$, and additional sources of error may be present that are unaccounted for. Assuming that the residuals from the model follow a Gaussian distribution $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2)$ with an a priori unknown variance, the variance-covariance matrix on the data $\mathbf{C}_\text{d}$ can be expressed proportional to the weight matrix:

$$\mathbf{C}_\text{d} = \sigma^2 \mathbf{W}^{-1} \tag{A.1.8}$$

Where the a priori variance on the residuals σ^2 can be approximated by the a posteriori variance of unit weight χ_ν^2, that is equivalent to the reduced chi-square of the weighted residuals between the data and the model (eq. (A.1.2)):

$$\sigma^2 \approx \chi_\nu^2 = \frac{\boldsymbol{\epsilon}^\text{T} \mathbf{W} \boldsymbol{\epsilon}}{\nu} \tag{A.1.9}$$

Where $\nu = n - m$ represents the number of degrees of freedom and is calculated from the number of observations n, minus the number of fitted model parameters m. The uncertainty on the model parameters $\mathbf{C}_{\tilde{\mathrm{m}}}$ can be propagated from the errors on the observations through the data covariance matrix $\mathbf{C}_{\mathrm{d}}$:

$$\mathbf{C}_{\tilde{\mathrm{m}}} = \mathbf{G}^{-\mathrm{g}}\mathbf{C}_{\mathrm{d}}\mathbf{G}^{-\mathrm{g}\mathrm{T}} \tag{A.1.10}$$

where $\mathbf{G}^{-\mathrm{g}}$ represents the generalised inverse operator that maps the data space to the model space in eq. (A.1.3). Filling in the generalised inverse into eq. (A.1.10) gives:

$$\mathbf{C}_{\tilde{\mathrm{m}}} = [\mathbf{G}^{\mathrm{T}}\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{W}\mathbf{C}_{\mathrm{d}}\mathbf{W}^{\mathrm{T}}\mathbf{G}[\mathbf{G}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{G}]^{-1} \tag{A.1.11}$$

When filling in eqs. (A.1.8) and (A.1.9) into eq. (A.1.11), this expression reduces to a more manageable form:

$$\mathbf{C}_{\tilde{\mathrm{m}}} = \chi^2_\nu[\mathbf{G}^{\mathrm{T}}\mathbf{W}\mathbf{G}]^{-1} \tag{A.1.12}$$

The standard deviation on each individual model parameter $\boldsymbol{\sigma}_{\tilde{\mathrm{m}}}$ (i.e., instrument drift parameters, gravity differences, and data tares) is recovered from the square root of the diagonal of $\mathbf{C}_{\tilde{\mathrm{m}}}$:

$$\boldsymbol{\sigma}_{\tilde{\mathrm{m}}} = \sqrt{\mathrm{diag}(\mathbf{C}_{\tilde{\mathrm{m}}})} \tag{A.1.13}$$

The difference in gravity between each benchmark and the anchor is a meaningful measure when this difference is compared between different campaigns. The mean change in relative gravity between the anchor and each benchmark between two campaigns can be calculated from the estimated model parameters recovered from eq. (A.1.3) and simply subtracted. The associated uncertainty on the change in gravity differences between the campaigns that comes from eq. (A.1.13) should be compounded:

$$\sigma_{\mathrm{i,campaign}_{2-1}} = \sqrt{\sigma^2_{\mathrm{i,campaign}_1} + \sigma^2_{\mathrm{i,campaign}_2}} \tag{A.1.14}$$

Giving the standard deviation on the gravity change between two campaigns for each benchmark. Circuits that were expressed relative to a substitute anchor had the uncertainties on the two relative solutions compounded using a similar expression to eq. (A.1.14) to find the benchmark uncertainties relative to network anchor.

Bibliography

Agnew, D. (2010), HARDISP, *IERS Conventions (2010), IERS Tech. Note*, *36*, 99–122.

Ahern, T., R. Casey, D. Barnes, R. Benson, T. Knight, and C. Trabant (2007), SEED Reference Manual, version 2.4, *PLACE: Incorporated Research Institutions for Seismology*.

Ahern, T., R. Benson, R. Casey, C. Trabant, and B. Weertman (2015), Improvements in Data Quality, Integration and Reliability: New Developments at the IRIS DMC, *Advances in Geosciences*, *40*, 31–35, doi:10.5194/adgeo-40-31-2015.

Altamimi, Z., P. Rebischung, L. Métivier, and X. Collilieux (2016), ITRF2014: A new release of the International Terrestrial Reference Frame modeling nonlinear station motions, *Journal of Geophysical Research: Solid Earth*, *121*(8), 6109–6131, doi:10.1002/2016JB013098.

Anastasiou, K., G. Hammond, D. Paul, K. Toland, A. Prasad, S. Bramsiepe, E. Passey, and H. Rakoczi ♡ (2022), An update to the development of the Wee-g: A high-sensitivity MEMS-based relative gravimeter for multi-pixel applications, *Tech. rep.*, Copernicus Meetings, doi: 10.5194/egusphere-egu22-4165.

Anderson, K. R., and M. P. Poland (2016), Bayesian estimation of magma supply, storage, and eruption rates using a multiphysical volcano model: Kīlauea Volcano, 2000 – 2012, *Earth and Planetary Science Letters*, *447*, 161–171, doi:10.1016/j.epsl.2016.04.029.

Anderson, K. R., M. P. Poland, J. H. Johnson, and A. Miklius (2015), Episodic deflation-inflation events at Kīlauea Volcano and implications for the shallow magma system, in *Hawaiian Volcanoes*, pp. 229–250, John Wiley & Sons, Inc, doi:10.1002/9781118872079.ch11.

Anderson, K. R., I. A. Johanson, M. R. Patrick, M. Gu, P. Segall, M. P. Poland, E. K. Montgomery-Brown, and A. Miklius (2019), Magma reservoir failure and the onset of caldera collapse at Kīlauea Volcano in 2018, *Science*, *366*(6470), eaaz1822, doi:10.1126/science.aaz1822.

Anderson, K. R., A. N, J. H. Foster, and N. Frazer (2020), Implications of deflation-inflation event models on Kīlauea Volcano, Hawaiʻi, *Journal of Volcanology and Geothermal Research*, *397*, 106,832, doi:10.1016/j.jvolgeores.2020.106832.

Anthony, R. E., A. T. Ringler, D. C. Wilson, M. Bahavar, and K. D. Koper (2020), How processing methodologies can distort and bias power spectral density estimates of seismic background noise, *Seismological Research Letters*, *91*(3), 1694–1706, doi:10.1785/0220190212.

Antoni-Micollier, L., D. Carbone, V. Ménoret, J. Lautier-Gaud, T. King, F. Greco, A. Messina, D. Contrafatto, and B. Desruelle (2022), Detecting volcano-related underground mass changes with a quantum gravimeter, *Earth and Space Science Open Archive*, doi:10.1029/2022GL097814.

Bagnardi, M., M. P. Poland, D. Carbone, S. Baker, M. Battaglia, and F. Amelung (2014), Gravity changes and deformation at Kīlauea Volcano, Hawaiʻi, associated with summit eruptive activity, 2009 – 2012, *Journal of Geophysical Research: Solid Earth*, *119*(9), 7288–7305, doi:10.1002/2014JB011506.

Battaglia, M., P. Segall, and C. Roberts (2003), The mechanics of unrest at Long Valley caldera, California. 2. Constraining the nature of the source using geodetic and micro-gravity data, *Journal of Volcanology and Geothermal Research*, *127*(3-4), 219–245, doi:10.1016/S0377-0273(03)00171-9.

Battaglia, M., J. Gottsmann, D. Carbone, and J. Fernández (2008), 4D volcano gravimetry, *Geophysics*, *73*(6), WA3–WA18, doi:10.1190/1.2977792.

Battaglia, M., M. P. Poland, and J. Kauahikaua (2012), GTOOLS: an interactive computer program to process gravity data for high-resolution applications, in *AGU Fall Meeting Abstracts*, vol. 2012, pp. GP43B–1143.

Battaglia, M., M. Lisowski, D. Dzurisin, M. P. Poland, S. Schilling, A. Diefenbach, and J. Wynn (2018), Mass addition at Mount St. Helens, Washington, inferred from repeated gravity surveys, *Journal of Geophysical Research: Solid Earth*, *123*(2), 1856–1874, doi:10.1002/2017JB014990.

Battaglia, M., A. C.-D. Patre, and A. F. Flinders (2022), gTOOLS, an open-source MATLAB program for processing high precision, relative gravity data for time-lapse gravity monitoring, *Computers & Geosciences*, *160*, 105,028, doi:10.1016/j.cageo.2021.105028.

Beilin, J. (2006), Apport de la gravimétrie absolue à la réalisation de la composante gravimétrique du Réseau Géodésique Français, *Institut Géographique National*.

Bemelmans, M. J., E. de Zeeuw-van Dalfsen, M. P. Poland, and I. A. Johanson (2021), Insight into the May 2015 summit inflation event at Kīlauea Volcano, Hawaiʻi, *Journal of Volcanology and Geothermal Research*, *415*, 107,250, doi:10.1016/j.jvolgeores.2021.107250.

Bennett, W. R. (1948), Spectra of quantized signals, *Bell System Technical Journal*, *27*(3), 446–472, doi:10.1002/j.1538-7305.1948.tb01340.x.

Berger, J., P. Davis, and G. Ekström (2004), Ambient Earth noise: a survey of the global seismographic network, *Journal of Geophysical Research: Solid Earth*, *109*(B11), doi:10.1029/2004JB003408.

Berrino, G. (1994), Gravity changes induced by height-mass variations at the Campi Flegrei caldera, *Journal of Volcanology and Geothermal Research*, *61*(3-4), 293–309, doi:10.1016/0377-0273(94)90010-8.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010), ObsPy: A Python toolbox for seismology, *Seismological Research Letters*, *81*(3), 530–533, doi:10.1785/gssrl.81.3.530.

Bormann, P., and E. Wielandt (2013), Seismic signals and noise, in *New Manual of Seismological Observatory Practice 2 (NMSOP2)*, pp. 1–62, Deutsches GeoForschungsZentrum GFZ, doi:10.2312/GFZ.NMSOP-2_ch4.

Bos, M., and H. Scherneck (2014), Onsala Space Observatory.

Brown, D., L. Ceranna, M. Prior, P. Mialle, and R. J. L. Bras (2012), The IDC seismic, hydroacoustic and infrasound global low and high noise models, *Pure and Applied Geophysics*, *171*(3-5), 361–375, doi:10.1007/s00024-012-0573-6.

Brown, G., S. Everett, H. Rymer, D. McGarvie, and I. Foster (1991), New light on caldera evolution—Askja, Iceland, *Geology*, *19*(4), 352–355, doi:10.1130/0091-7613(1991)019<0352:NLOCEA>2.3.CO;2.

Burtin, A., L. Bollinger, J. Vergne, R. Cattin, and J. Nábělek (2008), Spectral analysis of seismic noise induced by rivers: A new tool to monitor spatiotemporal changes in stream hydrodynamics, *Journal of Geophysical Research: Solid Earth*, *113*(B5), doi:10.1029/2007JB005034.

Butler, K. E., and R. D. Russell (1993), Subtraction of powerline harmonics from geophysical records, *Geophysics*, *58*(6), 898–903, doi:10.1190/1.1443474.

Butler, K. E., and R. D. Russell (2003), Cancellation of multiple harmonic noise series in geophysical records, *Geophysics*, *68*(3), 1083–1090, doi:10.1190/1.1581080.

Camacho, A. G., P. Vajda, C. A. Miller, and J. Fernández (2021), A free-geometry geodynamic modelling of surface gravity changes using Growth-dg software, *Scientific Reports*, *11*(1), 1–17, doi:10.1038/s41598-021-02769-z.

Camelbeeck, T., and T. van Eck (1994), The Roer Valley Graben earthquake of 13 April 1992 and its seismotectonic setting, *Terra Nova*, *6*(3), 291–300, doi:10.1111/j.1365-3121.1994.tb00499.x.

Carbone, D., and F. Greco (2007), Review of microgravity observations at Mt. Etna: a powerful tool to monitor and study active volcanoes, *Pure and Applied Geophysics*, *164*(4), 769–790, doi:10.1007/s00024-007-0194-7.

Carbone, D., and M. P. Poland (2012), Gravity fluctuations induced by magma convection at Kīlauea Volcano, Hawai‘i, *Geology*, *40*(9), 803–806, doi:10.1130/G33060.1.

Carbone, D., M. P. Poland, M. R. Patrick, and T. R. Orr (2013), Continuous gravity measurements reveal a low-density lava lake at Kīlauea Volcano, Hawai‘i, *Earth and Planetary Science Letters*, *376*, 178–185, doi:10.1016/j.epsl.2013.06.024.

Carbone, D., L. Zuccarello, A. Messina, S. Scollo, and H. Rymer (2015), Balancing bulk gas accumulation and gas output before and during lava fountaining episodes at Mt. Etna, *Scientific reports*, *5*(1), 1–11, doi:10.1038/srep18049.

Carbone, D., M. P. Poland, M. Diament, and F. Greco (2017), The added value of time-variable microgravimetry to the understanding of how volcanoes work, *Earth-Science Reviews*, *169*, 146–179, doi:10.1016/j.earscirev.2017.04.014.

Carbone, D., F. Cannavò, F. Greco, R. Reineman, and R. J. Warburton (2019), The benefits of using a network of superconducting gravimeters to monitor and study active volcanoes, *Journal of Geophysical Research: Solid Earth*, *124*(4), 4035–4050, doi:10.1029/2018JB017204.

Carbone, D., L. Antoni-Micollier, G. Hammond, E. de Zeeuw-van Dalfsen, E. Rivalta, C. Bonadonna, A. Messina, J. Lautier-Gaud, K. Toland, M. Koymans, et al. (2020), The NEWTON-g gravity imager: Toward new paradigms for terrain gravimetry, *Frontiers in Earth Science*, p. 452, doi:10.3389/feart.2020.573396.

Castellaro, S., and F. Mulargia (2012), A statistical low noise model of the Earth, *Seismological Research Letters*, *83*(1), 39–48, doi:10.1785/gssrl.83.1.39.

Caswell, T. A., M. Droettboom, A. Lee, J. Hunter, E. Firing, D. Stansby, J. Klymak, T. Hoffmann, E. S. de Andrade, N. Varoquaux, J. H. Nielsen, B. Root, P. Elson, R. May, D. Dale, J.-J. Lee, J. K. Seppänen, D. McDougall, A. Straw, P. Hobson, C. Gohlke, T. S. Yu, E. Ma, A. F. Vincent, S. Silvester, C. Moad, N. Kniazev, P. Ivanov, E. Ernest, and J. Katins (2020), matplotlib/matplotlib: Rel: v3.2.1, doi:10.5281/zenodo.3714460.

Cattin, R., S. Mazzotti, and L.-M. Baratin (2015), GravProcess: An easy-to-use MATLAB software to process campaign gravity data and evaluate the associated uncertainties, *Computers & Geosciences*, *81*, 20–27, doi:10.1016/j.cageo.2015.04.005.

Cauzzi, C., and J. Clinton (2013), A high- and low-noise model for high-quality strong-motion accelerometer stations, *Earthquake Spectra*, *29*(1), 85–102, doi:10.1193/1.4000107.

Cohen, M. B., R. Said, and U. Inan (2010), Mitigation of 50–60 Hz power line interference in geophysical data, *Radio Science*, *45*(06), 1–12, doi:10.1029/2010RS004420.

Cooley, J. W., and J. W. Tukey (1965), An algorithm for the machine calculation of complex Fourier series, *Mathematics of Computation*, *19*(90), 297–301.

Cooper, A. H. (2008), The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings – an automated approach, *Journal of the Audio Engineering Society*.

Cooper, A. J. (2010), An automated approach to the electric network frequency (ENF) criterion - theory and practice, *International Journal of Speech Language and the Law*, *16*(2), doi:10.1558/ijsll.v16i2.193.

Coyne, J., D. Bobrov, P. Bormann, E. Duran, P. Grenard, G. Haralabus, I. Kitov, and Y. Starovoit (2012), CTBTO: Goals, networks, data analysis and data availability, in *New Manual of Seismological Observatory Practice 2 (NMSOP-2)*, pp. 1–41, Deutsches GeoForschungsZentrum GFZ, doi:10.2312/GFZ.NMSOP-2_ch15.

Crossley, D., J. Hinderer, and U. Riccardi (2013), The measurement of surface gravity, *Reports on Progress in Physics*, *76*(4), 046,101, doi:10.1088/0034-4885/76/4/046101.

Crozier, J., and L. Karlstrom (2021), Wavelet-Based Characterization of Very-Long-Period Seismicity Reveals Temporal Evolution of Shallow Magma System Over the 2008 – 2018 Eruption of Kīlauea Volcano, *Journal of Geophysical Research: Solid Earth*, *126*(6), e2020JB020,837, doi:10.1029/2020JB020837.

de Zeeuw-van Dalfsen, E., H. Rymer, F. Sigmundsson, and E. Sturkell (2005), Net gravity decrease at Askja volcano, Iceland: constraints on processes responsible for continuous caldera deflation, 1988–2003, *Journal of Volcanology and Geothermal Research*, *139*(3-4), 227–239, doi:10.1016/j.jvolgeores.2004.08.008.

de Zeeuw-van Dalfsen, E., R. Pedersen, A. Hooper, and F. Sigmundsson (2012), Subsidence of Askja caldera 2000–2009: Modelling of deformation processes at an extensional plate boundary, constrained by time series InSAR analysis, *Journal of Volcanology and Geothermal Research*, *213*, 72–82, doi:10.1016/j.jvolgeores.2011.11.004.

de Zeeuw-van Dalfsen, E., H. Rymer, E. Sturkell, R. Pedersen, A. Hooper, F. Sigmundsson, and B. Ófeigsson (2013), Geodetic data shed light on ongoing caldera subsidence at Askja, Iceland, *Bulletin of Volcanology*, *75*(5), 1–13, doi:10.1007/s00445-013-0709-2.

De Zeeuw-van Dalfsen, E., R. Sleeman, et al. (2018), A Permanent, Real-Time Monitoring Network for the Volcanoes Mount Scenery and The Quill in the Caribbean Netherlands, *Geosciences*, *8*(9), 320, doi:10.3390/geosciences8090320.

Dehant, V., P. Defraigne, and J. Wahr (1999), Tides for a convective Earth, *Journal of Geophysical Research: Solid Earth*, *104*(B1), 1035–1058, doi:10.1029/1998JB900051.

Dost, B. (1994), The ORFEUS Data Center, *Annals of Geophysics*, *37*(5).

Dost, B., and H. Haak (2002), *A comprehensive description of the KNMI seismological instrumentation*, Royal Netherlands Meteorological Institute.

Dougherty, S. L., E. S. Cochran, and R. M. Harrington (2019), The large-N seismic survey in Oklahoma (LASSO) experiment, *Seismological Research Letters*, *90*(5), 2051–2057, doi:10.1785/0220190094.

Drouin, V., F. Sigmundsson, B. G. Ófeigsson, S. Hreinsdóttir, E. Sturkell, and P. Einarsson (2017), Deformation in the Northern Volcanic Zone of Iceland 2008–2014: An interplay of tectonic, magmatic, and glacial isostatic deformation, *Journal of Geophysical Research: Solid Earth*, *122*(4), 3158–3178, doi:10.1002/2016JB013206.

Dvorak, J. J., and D. Dzurisin (1993), Variations in magma supply rate at Kīlauea Volcano, Hawai‘i, *Journal of Geophysical Research: Solid Earth*, *98*(B12), 22,255–22,268, doi:10.1029/93jb02765.

Dzurisin, D. (2003), A comprehensive approach to monitoring volcano deformation as a window on the eruption cycle, *Reviews of Geophysics*, *41*(1), doi:10.1029/2001RG000107.

Dzurisin, D., and M. P. Poland (2019), Magma supply to Kīlauea volcano, Hawai‘i, from inception to now: Historical perspective, current state of knowledge, and future challenges, in *Field Volcanology: A Tribute to the Distinguished Career of Don Swanson*, Geological Society of America, doi:10.1130/2018.2538(12).

Dzurisin, D., L. A. Anderson Kyle R., G. P. Eaton, R. Y. Koyanagi, P. W. Lipman, J. P. Lockwood, R. T. Okamura, G. S. Puniwai, M. K. Sako, and K. M. Yamashita (1980), Geophysical observations of Kīlauea Volcano, Hawai‘i, 2. Constraints on the magma supply during November 1975 – September 1977, *Journal of Volcanology and Geothermal Research*, *7*(3-4), 241–269.

Egbert, G. D., and S. Y. Erofeeva (2002), Efficient inverse modeling of barotropic ocean tides, *Journal of Atmospheric and Oceanic technology*, *19*(2), 183–204, doi:10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2.

Eichelberger, J. (1980), Vesiculation of mafic magma during replenishment of silicic magma reservoirs, *Nature*, *288*(5790), 446–450, doi:10.1038/288446a0.

Einarsson, P., and B. Brandsdóttir (2021), Seismicity of the Northern Volcanic Zone of Iceland, *Frontiers in Earth Science*, *9*, 628,967, doi:10.3389/feart.2021.628967.

Elias, T., C. Kern, K. A. Horton, A. J. Sutton, and H. Garbeil (2018), Measuring SO_2 emission rates at Kīlauea Volcano, Hawai‘i, using an array of upward-looking UV spectrometers, 2014 – 2017, *Frontiers in Earth Science*, *6*, 214, doi:10.3389/feart.2018.00214.

Elósegui, P., J. Davis, D. Oberlander, R. Baena, and G. Ekström (2006), Accuracy of high-rate GPS for seismology, *Geophysical Research Letters*, *33*(11), doi:10.1029/2006GL026065.

Evers, L. G., L. Ceranna, H. W. Haak, A. Le Pichon, and R. W. Whitaker (2007), A seismoacoustic analysis of the gas-pipeline explosion near Ghislenghien in Belgium, *Bulletin of the Seismological Society of America*, *97*(2), 417–425, doi:10.1785/0120060061.

Flinders, A. F., M. P. Poland, S. Conway, M. Bagnardi, M. Battaglia, S. Brantley, I. Johanson, A. Mikilius, A. Pitty, P. MacQueen, and N. Young (2022), Data of the Hawaiian Volcano Observatory Kīlauea Campaign Gravity Network (KCGN) (2009 – 2017), doi:doi.org/10.5066/P9INL9SG.

Forbriger, T. (2007), Reducing magnetic field induced noise in broad-band seismic recordings, *Geophysical Journal International*, *169*(1), 240–258, doi:10.1111/j.1365-246X.2006.03295.x.

Fowler, C. M. R., C. M. R. Fowler, and M. Fowler (1990), *The Solid Earth: an introduction to global geophysics*, Cambridge University Press.

Francis, O. (2021), Performance assessment of the relative gravimeter Scintrex CG-6, *Journal of Geodesy*, *95*(10), 1–14, doi:10.1007/s00190-021-01572-y.

Francis, O., C. Lampitelli, G. Klein, M. Van Camp, and V. Palinkas (2011), Comparison between the transfer functions of three superconducting gravimeters, *Marées Terrestres Bulletin d'Informations*, *147*, 11,857–11,868.

Fricke, J., L. Evers, P. Smets, K. Wapenaar, and D. Simons (2014), Infrasonic interferometry applied to microbaroms observed at the Large Aperture Infrasound Array in the Netherlands, *Journal of Geophysical Research: Atmospheres*, *119*(16), 9654–9665, doi:10.1002/2014JD021663.

Furuya, M., S. Okubo, W. Sun, Y. Tanaka, J. Oikawa, H. Watanabe, and T. Maekawa (2003), Spatiotemporal gravity changes at Miyakejima Volcano, Japan: Caldera collapse, explosive eruptions and magma movement, *Journal of Geophysical Research: Solid Earth*, *108*(B4), doi: 10.1029/2002JB001989.

Gange, M. (2011), Low-frequency and tonal characteristics of Transformer noise, in *Proceedings of Acoustics*.

García, A., R. Ortiz, J. M. Marrero, N. Sánchez, J. Vila, A. M. Correig, R. Marcià, R. Sleeman, and M. Tárraga (2006), Monitoring the reawakening of Canary Islands' Teide volcano, *Eos, Transactions American Geophysical Union*, *87*(6), 61–65, doi:10.1029/2006EO060001.

Garg, R., A. L. Varna, and M. Wu (2011), "Seeing" ENF: Natural time stamp for digital video via optical sensing and signal processing, in *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, p. 23–32, Association for Computing Machinery, doi:10.1145/2072298.2072303.

Garg, R., A. L. Varna, and M. Wu (2012), Modeling and analysis of electric network frequency signal for timestamp verification, in *2012 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 67–72, IEEE, doi:10.1109/WIFS.2012.6412627.

Giniaux, J. (2021), Gravity campaigns at Askja Volcano (Iceland), in 2015, 2016 and 2017, using the Scintrex CG-5 968, doi:10.5518/1030.

Giniaux, J. M. (2019), Constraints on Askja Volcano, Iceland, from Surface Deformation and Gravity Change.

Giniaux, J. M., A. J. Hooper, and M. Bagnardi (2017), New Methods for Assessing and Reducing Uncertainty in Microgravity Studies, in *AGU Fall Meeting Abstracts*, vol. 2017, pp. G54A–06, doi:10.1089/space.2020.0044.

Gorjão, L. R., R. Jumar, H. Maass, V. Hagenmeyer, G. C. Yalcin, J. Kruse, M. Timme, C. Beck, D. Witthaut, and B. Schäfer (2020), Open database analysis of scaling and spatio-temporal properties of power grid frequencies, *Nature Communications*, *11*(1), 1–11, doi: 10.1038/s41467-020-19732-7.

Gottsmann, J., and H. Rymer (2002), Deflation during caldera unrest: constraints on subsurface processes and hazard prediction from gravity–height data, *Bulletin of Volcanology*, *64*(5), 338–348, doi:10.1007/s00445-002-0212-7.

Gottsmann, J., G. Berrino, H. Rymer, and G. Williams-Jones (2003), Hazard assessment during caldera unrest at the Campi Flegrei, Italy: a contribution from gravity–height gradients, *Earth and Planetary Science Letters*, *211*(3-4), 295–309, doi:10.1016/S0012-821X(03)00225-5.

Grapenthin, R., Y. Cheng, M. Angarita, D. Tan, F. J. Meyer, D. Fee, and A. Wech (2022), Return from dormancy: Rapid inflation and seismic unrest driven by transcrustal magma transfer at Mt. Edgecumbe (L'úx Shaa) volcano, Alaska, *Geophysical Research Letters*, *49*(20), e2022GL099,464, doi:10.1029/2022GL099464.

Greenfield, T., R. S. White, and S. Roecker (2016), The magmatic plumbing system of the Askja central volcano, Iceland, as imaged by seismic tomography, *Journal of Geophysical Research: Solid Earth*, *121*(10), 7211–7229, doi:10.1002/2016JB013163.

Greenfield, T., R. S. White, T. Winder, and T. Ágústsdóttir (2020), Seismicity of the Askja and Bárðarbunga volcanic systems of Iceland, 2009–2015, *Journal of Volcanology and Geothermal Research*, *391*, 106,432, doi:10.1016/j.jvolgeores.2018.08.010.

Gudmundsson, M. T., K. Jónsdóttir, A. Hooper, E. P. Holohan, S. A. Halldórsson, B. G. Ófeigsson, S. Cesca, K. S. Vogfjörd, F. Sigmundsson, T. Högnadóttir, et al. (2016), Gradual caldera collapse at Bárdarbunga volcano, Iceland, regulated by lateral magma outflow, *Science*, *353*(6296), aaf8988, doi:10.1126/science.aaf8988.

Halldórsdóttir, S., H. Björnsson, A. K. Mortensen, G. Axelsson, and Á. Guðmundsson (2010), Temperature model and volumetric assessment of the Krafla geothermal field in N-Iceland, in *World Geothermal Congress*, pp. 25–29.

Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant (2020), Array programming with NumPy, *Nature*, *585*(7825), 357–362, doi:10.1038/s41586-020-2649-2.

Hartley, M., and T. Thordarson (2012), Formation of Öskjuvatn caldera at Askja, North Iceland: Mechanism of caldera collapse and implications for the lateral flow hypothesis, *Journal of Volcanology and Geothermal Research*, *227*, 85–101, doi:10.1016/j.jvolgeores.2012.02.009.

Havskov, J., and G. Alguacil (2016), *Instrumentation in Earthquake Seismology*, Springer International Publishing, doi:10.1007/978-3-319-21314-9.

Hector, B., and J. Hinderer (2016), pyGrav, a Python-based program for handling and processing relative gravity data, *Computers & Geosciences*, *91*, 90–97, doi:10.1016/j.cageo.2016.03.010.

Heimisson, E. R., P. Einarsson, F. Sigmundsson, and B. Brandsdóttir (2015), Kilometer-scale Kaiser effect identified in Krafla volcano, Iceland, *Geophysical Research Letters*, *42*(19), 7958–7965, doi:10.1002/2015GL065680.

Heinzel, G., A. Rüdiger, and R. Schilling (2002), Spectrum and spectral density estimation by the discrete Fourier transform (DFT), including a comprehensive list of window functions and some new flat-top windows, *Max Plank Inst*, *12*.

Herring, T., R. King, S. McClusky, et al. (2010), Introduction to GAMIT/GLOBK, *Massachusetts Institute of Technology, Cambridge, Massachusetts*.

Hinderer, J., B. Hector, A. Mémin, and M. Calvo (2016), Hybrid gravimetry as a tool to monitor surface and underground mass changes, in *International Symposium on Earth and Environmental Sciences for Future Generations*, pp. 123–130, Springer, doi:10.1007/1345_2016_253.

Houghton, B., D. Swanson, R. Carey, J. Rausch, and A. Sutton (2011), Pigeonholing pyroclasts: Insights from the 19 March 2008 explosive eruption of Kīlauea Volcano, *Geology*, *39*(3), 263–266, doi:10.1130/g31509.1.

Hunter, J. D. (2007), Matplotlib: a 2D graphics environment, *Computing in Science & Engineering*, *9*(3), 90–95, doi:10.1109/MCSE.2007.55.

Hwang, C., C.-G. Wang, and L.-H. Lee (2002), Adjustment of relative gravity measurements using weighted and datum-free constraints, *Computers & Geosciences*, *28*(9), 1005–1015, doi:10.1016/S0098-3004(02)00005-5.

Imanishi, Y., K. Nawa, Y. Tamura, and H. Ikeda (2022), Effects of vertical nonlinearity on the superconducting gravimeter CT #036 at Ishigakijima, Japan, *Earth, Planets and Space*, *74*(1), 1–15, doi:10.1186/s40623-022-01609-2.

Jachens, R. C., and G. P. Eaton (1980), Geophysical observations of Kīlauea Volcano, Hawai'i, 1. Temporal gravity variations related to the 29 November, 1975, M = 7.2 earthquake and associated summit collapse, *Journal of Volcanology and Geothermal Research*, *7*(3-4), 225–240.

Jakobsson, S. P., K. Jónasson, and I. A. Sigurdsson (2008), The three igneous rock series of Iceland, *Jökull*, *58*, 117–138.

Jo, M.-J., H.-S. Jung, and J.-S. Won (2015), Detecting the source location of recent summit inflation via three-dimensional InSAR observation of Kīlauea Volcano, *Remote Sensing*, *7*(11), 14,386–14,402, doi:10.3390/rs71114386.

Johannesson, H., and K. Sæmundsson (2009), Geological Map of Iceland, 1:500 000, Tectonics.

Johnson, D. J. (1992), Dynamics of magma storage in the summit reservoir of Kīlauea Volcano, Hawai'i, *Journal of Geophysical Research: Solid Earth*, *97*(B2), 1807–1820, doi:10.1029/91JB02839.

Johnson, D. J., A. A. Eggers, M. Bagnardi, M. Battaglia, M. P. Poland, and A. Miklius (2010), Shallow magma accumulation at Kīlauea Volcano, Hawai'i, revealed by microgravity surveys, *Geology*, *38*(12), 1139–1142, doi:10.1130/G31323.1.

Jousset, P., H. Mori, and H. Okada (2003), Elastic models for the magma intrusion associated with the 2000 eruption of Usu Volcano, Hokkaido, Japan, *Journal of Volcanology and Geothermal Research*, *125*(1-2), 81–106, doi:10.1016/S0377-0273(03)00090-8.

Jurkevics, A. (1988), Polarization analysis of three-component array data, *Bulletin of the seismological society of America*, *78*(5), 1725–1743, doi:10.1785/BSSA0780051725.

Kauahikaua, J. (1993), Geophysical characteristics of the hydrothermal systems of Kīlauea Volcano, Hawai'i, *Geothermics*, *22*(4), 271–299, doi:10.1016/0375-6505(93)90004-7.

Kauahikaua, J., and A. Miklius (2003), Long-term trends in microgravity at Kīlauea's summit during the Pu'u Ō'ō-Kūpaianaha eruption, *US Geological Survey Professional Paper*, *1676*(1676), 165.

Kennedy, J. R. (2021), GSadjust, doi:10.5066/P9YEIOU8.

Kennett, B., and E. Engdahl (1991), Traveltimes for global earthquake location and phase identification, *Geophysical Journal International*, *105*(2), 429–465, doi:10.1111/j.1365-246X.1991.tb06724.x.

Kimman, W., X. Campman, and J. Trampert (2012), Characteristics of seismic noise: fundamental and higher mode energy observed in the northeast of the Netherlands, *Bulletin of the Seismological Society of America*, *102*(4), 1388–1399, doi:10.1785/0120110069.

Klun, M., D. Zupan, J. Lopatič, and A. Kryžanowski (2019), On the application of laser vibrometry to perform structural health monitoring in non-stationary conditions of a hydropower dam, *Sensors*, *19*(17), 3811, doi:10.3390/s19173811.

KNMI (1993), Netherlands Seismic and Acoustic Network, Royal Netherlands Meteorological Institute (KNMI), Other/Seismic Network, 10.21944/e970fd34-23b9-3411-b366-e4f72877d2c5.

KNMI (2006), Caribbean Netherlands Seismic Network, Royal Netherlands Meteorological Institute (KNMI), Other/Seismic Network, doi:10.21944/dffa7a3f-7e3a-3b33-a436-516a01b6af3f.

Koymans, M., J. Domingo Ballesta, E. Ruigrok, R. Sleeman, L. Trani, and L. Evers (2021), Performance assessment of geophysical instrumentation through the automated analysis of power spectral density estimates, *Earth and Space Science*, *8*(9), e2021EA001,675, doi:10.1029/2021EA001675.

Koymans, M., E. de Zeeuw-van Dalfsen, L. Evers, and M. Poland (2022), Microgravity change during the 2008–2018 Kīlauea summit eruption: nearly a decade of subsurface mass accumulation, *Journal of Geophysical Research: Solid Earth*, *127*(9), e2022JB024,739, doi:10.1029/2022JB024739.

Koymans, M., E. de Zeeuw-van Dalfsen, A. Hooper, and J. Araya (2023), Dataset of microgravity campaign measurements from the Askja volcano, doi:10.4121/912B666B-95C0-4F93-9485-E98256517991.V1.

Koymans, M. R. (2022a), Web application for microgravity campaign analysis, doi:10.5281/zenodo.6466389.

Koymans, M. R. (2022b), NEWTON-g/AQG2MSEED: Version 1.0.0, doi:10.5281/zenodo.7346903.

Koymans, M. R. (2022c), NEWTON-g/WEEG2MSEED: Version 1.0.0, doi:10.5281/zenodo.7346927.

Křen, P., V. Pálinkáš, M. Val'ko, and P. Mašika (2021), Improved measurement model for FG5/X gravimeters, *Measurement*, *171*, 108,739, doi:10.1016/j.measurement.2020.108739.

Lampert, T. A., and S. E. O'Keefe (2010), A survey of spectrogram track detection algorithms, *Applied Acoustics*, *71*(2), 87–100, doi:10.1016/j.apacoust.2009.08.007.

Lecocq, T., S. P. Hicks, K. V. Noten, K. van Wijk, P. Koelemeijer, R. S. M. D. Plaen, F. Massin, G. Hillers, R. E. Anthony, M.-T. Apoloner, M. Arroyo-Solórzano, J. D. Assink, P. Büyükakpınar, A. Cannata, F. Cannavo, S. Carrasco, C. Caudron, E. J. Chaves, D. G. Cornwell, D. Craig, O. F. C. den Ouden, J. Diaz, S. Donner, C. P. Evangelidis, L. Evers, B. Fauville, G. A. Fernandez, D. Giannopoulos, S. J. Gibbons, T. Girona, B. Grecu, M. Grunberg, G. survey of spectrogram track detection algorithmsörgy Hetényi, A. Horleston, A. Inza, J. C. E. Irving, M. Jamalreyhani, A. Kafka, M. R. Koymans, C. R. Labedz, E. Larose, N. J. Lindsey, M. McKinnon, T. Megies, M. S. Miller, W. Minarik, L. Moresi, V. H. Márquez-Ramírez, M. Möllhoff, I. M. Nesbitt, S. Niyogi, J. Ojeda, A. Oth, S. Proud, J. Pulli, L. Retailleau, A. E. Rintamäki, C. Satriano, M. K. Savage, S. Shani-Kadmiel, R. Sleeman, E. Sokos, K. Stammler, A. E. Stott, S. Subedi, M. B. Sørensen, T. Taira, M. Tapia, F. Turhan, B. van der Pluijm, M. Vanstone, J. Vergne, T. A. T. Vuorinen, T. Warren, J. Wassermann, and H. Xiao (2020), Global quieting of high-frequency seismic noise due to COVID-19 pandemic lockdown measures, *Science*, *369*(6509), 1338–1343, doi:10.1126/science.abd2438.

Levkov, C., G. Mihov, R. Ivanov, I. Daskalov, I. Christov, and I. Dotsinsky (2005), Removal of power-line interference from the ECG: a review of the subtraction procedure, *BioMedical Engineering OnLine*, *4*(1), 1–18, doi:10.1186/1475-925X-4-50.

Li, Z., Z. Peng, D. Hollis, L. Zhu, and J. McClellan (2018), High-resolution seismic event detection using local similarity for large-N arrays, *Scientific reports*, *8*(1), 1646, doi:10.1038/s41598-018-19728-w.

Loeb, L. B. (1965), *Electrical coronas, their basic physical mechanisms*, Univ of California Press.

Longman, I. (1959), Formulas for computing the tidal accelerations due to the moon and the sun, *Journal of Geophysical Research*, *64*(12), 2351–2355, doi:10.1029/JZ064i012p02351.

Lundgren, P., M. P. Poland, A. Miklius, T. Orr, S.-H. Yun, E. Fielding, Z. Liu, A. Tanaka, W. Szeliga, S. Hensley, et al. (2013), Evolution of dike opening during the March 2011 Kamoamoa fissure eruption, Kīlauea Volcano, Hawai'i, *Journal of Geophysical Research: Solid Earth*, *118*(3), 897–914, doi:10.1002/jgrb.50108.

Lyard, F., F. Lefevre, T. Letellier, and O. Francis (2006), Modelling the global ocean tides: modern insights from FES2004, *Ocean Dynamics*, *56*(5), 394–415, doi:10.1007/s10236-006-0086-x.

McCubbine, J., F. C. Tontini, V. Stagpoole, E. Smith, and G. O'Brien (2018), Gsolve, a Python computer program with a graphical user interface to transform relative gravity survey measurements to absolute gravity values and gravity anomalies, *SoftwareX*, *7*, 129–137, doi:10.1016/j.softx.2018.04.003.

McNamara, and R. P. Buland (2004), Ambient noise levels in the continental United States, *Bulletin of the Seismological Society of America*, *94*(4), 1517–1527, doi:10.1785/012003001.

McNamara, D., and R. Boaz (2006a), PQLX: A software tool to evaluate seismic station performance, in *AGU Fall Meeting Abstracts*.

McNamara, D., C. Hutt, L. Gee, H. M. Benz, and R. Buland (2009), A method to establish seismic noise baselines for automated station assessment, *Seismological Research Letters*, *80*(4), 628–637, doi:10.1785/gssrl.80.4.628.

McNamara, D. E., and R. Boaz (2006b), *Seismic noise analysis system using power spectral density probability density functions: A stand-alone software package*, Citeseer.

Menke, W. (2018), *Geophysical Data Analysis: Discrete Inverse Theory*, Academic press.

Ménoret, V., P. Vermeulen, N. Le Moigne, S. Bonvalot, P. Bouyer, A. Landragin, and B. Desruelle (2018), Gravity measurements below 10^{-9} g with a transportable absolute quantum gravimeter, *Scientific reports*, *8*(1), 1–11, doi:10.1038/s41598-018-30608-1.

Mentink, J. H., and L. G. Evers (2011), Frequency response and design parameters for differential microbarometers, *The Journal of the Acoustical Society of America*, *130*(1), 33–41, doi:10.1121/1.3596718.

Middlemiss, R., A. Samarelli, D. Paul, J. Hough, S. Rowan, and G. Hammond (2016), Measurement of the Earth tides with a MEMS gravimeter, *Nature*, *531*(7596), 614–617, doi:10.1038/nature17397.

Miller, C., H. Le Mével, G. Currenti, G. Williams-Jones, and B. Tikoff (2017), Microgravity changes at the Laguna del Maule volcanic field: Magma-induced stress changes facilitate mass addition, *Journal of Geophysical Research: Solid Earth*, *122*(4), 3179–3196, doi:10.1002/2017JB014048.

Miller, C. A., G. Currenti, I. Hamling, and G. Williams-Jones (2018), Mass transfer processes in a post eruption hydrothermal system: parameterisation of microgravity changes at Te Maari craters, New Zealand, *Journal of Volcanology and Geothermal Research*, *357*, 39–55, doi:10.1016/j.jvolgeores.2018.04.005.

Mitchell, M. A., R. S. White, S. Roecker, and T. Greenfield (2013), Tomographic image of melt storage beneath Askja Volcano, Iceland using local microseismicity, *Geophysical Research Letters*, *40*(19), 5040–5046, doi:10.1002/grl.50899.

Mogi, K. (1958), Relations between the eruptions of various volcanoes and the deformations of the ground surfaces around them, *Bulletin of the Earthquake Research Institute*, *36*, 99–134.

Murray, A. S., and R. M. Tracey (2001), Best practice in gravity surveying, *Geoscience Australia*, *3*(5).

National Land Survey of Iceland (2020), Iceland Hillshade Digital Elevation Model (50m).

Neal, C. A., S. Brantley, L. Antolik, J. Babb, M. Burgess, K. Calles, M. Cappos, J. Chang, S. Conway, L. Desmither, et al. (2019), The 2018 rift eruption and summit collapse of Kīlauea Volcano, *Science*, *363*(6425), 367–374, doi:10.1126/science.aav7046.

Network Of Superconducting Gravimeters (1997), International Geodynamics and Earth Tide Service, doi:10.7914/SN/SG.

Niebauer, T., G. Sasagawa, J. E. Faller, R. Hilt, and F. Klopping (1995), A new generation of absolute gravimeters, *Metrologia*, *32*(3), 159, doi:10.1088/0026-1394/32/3/004.

Nikkhoo, M., and E. Rivalta (2022), Analytical solutions for gravity changes caused by triaxial volumetric sources, *Geophysical Research Letters*, *49*(8), e2021GL095,442, doi:10.1029/2021GL095442.

Nikkhoo, M., E. Rivalta, and D. Carbone (2018), Parameters definition for devices design, doi: 10.5281/zenodo.1492735.

Náttúrufræðistofnun Íslands – Icelandic Institute of Natural History (2018), NI_VG25 Vistgerðir á Íslandi 2. útgáfa – 1:25.000.

Okubo, S. (1991), Potential and gravity changes raised by point dislocations, *Geophysical Journal International*, *105*(3), 573–586, doi:10.1111/j.1365-246X.1991.tb00797.x.

Olsson, P.-A., G. Milne, H.-G. Scherneck, and J. Ågren (2015), The relation between gravity rate of change and vertical displacement in previously glaciated areas, *Journal of Geodynamics*, *83*, 76–84, doi:10.1016/j.jog.2014.09.011.

OpenStreetMap contributors (2017), Planet dump retrieved from https://planet.osm.org.

Oppenheim, A. V., and R. W. Schafer (2009), *Discrete-Time Signal Processing*, 3rd ed., Prentice Hall Press, USA.

Orr, T. R., M. P. Poland, M. R. Patrick, W. A. Thelen, A. J. Sutton, T. Elias, C. R. Thornber, C. Parcheta, and K. M. Wooten (2015), Kīlauea 5 – 9 March 2011 Kamoamoa fissure eruption and its relation to 30+ years of activity from Puʻu ʻŌʻō, *Hawaiian Volcanoes: From Source to Surface*, *208*, 393, doi:10.1002/9781118872079.ch18.

Pagli, C., F. Sigmundsson, T. Arnadóttir, P. Einarsson, and E. Sturkell (2006), Deflation of the Askja volcanic system: constraints on the deformation source from combined inversion of satellite radar interferograms and GPS measurements, *Journal of Volcanology and Geothermal Research*, *152*(1-2), 97–108, doi:10.1016/j.jvolgeores.2005.09.014.

Parks, M., B. Ófeigsson, V. Drouin, F. Sigmundsson, A. Hooper, H. Geirsson, S. Hreinsdóttir, H. Friðriksdóttir, E. Sturkell, Á. Hjartadóttir, et al. (2022), Deformation observations and geodetic modelling during the recent unrest at Askja volcano, *Tech. rep.*, Copernicus Meetings, doi:10.5194/egusphere-egu22-9381.

Patrick, M. R., T. Orr, A. Sutton, E. Lev, W. Thelen, and D. Fee (2016), Shallowly driven fluctuations in lava lake outgassing (gas pistoning), Kīlauea Volcano, *Earth and Planetary Science Letters*, *433*, 326–338, doi:10.1016/j.epsl.2015.10.052.

Patrick, M. R., T. Orr, K. R. Anderson, and D. Swanson (2019a), Eruptions in sync: Improved constraints on Kīlauea Volcano's hydraulic connection, *Earth and Planetary Science Letters*, *507*, 50–61, doi:10.1016/j.epsl.2018.11.030.

Patrick, M. R., D. Swanson, and T. Orr (2019b), A review of controls on lava lake level: insights from Halemaʻumaʻu Crater, Kīlauea Volcano, *Bulletin of Volcanology*, *81*(3), 13, doi:10.1007/s00445-019-1268-y.

Patrick, M. R., B. F. Houghton, K. R. Anderson, M. P. Poland, E. Montgomery-Brown, I. Johanson, W. Thelen, and T. Elias (2020), The cascading origin of the 2018 Kīlauea eruption and implications for future forecasting, *Nature Communications*, *11*(1), doi:10.1038/s41467-020-19190-1.

Pedersen, G., A. Höskuldsson, T. Dürig, T. Thordarson, I. Jonsdottir, M. Riishuus, B. Óskarsson, S. Dumont, E. Magnússon, M. T. Gudmundsson, et al. (2017), Lava field evolution and emplacement dynamics of the 2014–2015 basaltic fissure eruption at Holuhraun, Iceland, *Journal of Volcanology and Geothermal Research*, *340*, 155–169, doi:10.1016/j.jvolgeores.2017.02.027.

Pedersen, H. A., N. Leroy, D. Zigone, M. Vallée, A. T. Ringler, and D. C. Wilson (2020), Using component ratios to detect metadata and instrument problems of seismic stations: Examples from 18 yr of GEOSCOPE data, *Seismological Research Letters*, *91*(1), 272–286, doi:10.1785/0220190180.

Petersen, G. M., S. Cesca, and M. K. and (2019), Automated Quality Control for Large Seismic Networks: Implementation and Application to the AlpArray Seismic Network, *Seismological Research Letters*, *90*(3), 1177–1190, doi:10.1785/0220180342.

Peterson, J. R. (1993), Observations and modeling of seismic background noise, doi:10.3133/ofr93322.

Poland, M. P. (2014), Time-averaged discharge rate of subaerial lava at Kīlauea Volcano, Hawai‘i, measured from TanDEM-X interferometry: Implications for magma supply and storage during 2011 – 2013, *Journal of Geophysical Research: Solid Earth*, *119*(7), 5464–5481, doi:10.1002/2014JB011132.

Poland, M. P., and D. Carbone (2016), Insights into shallow magmatic processes at Kīlauea Volcano, Hawai‘i, from a multiyear continuous gravity time series, *Journal of Geophysical Research: Solid Earth*, *121*(7), 5477–5492, doi:10.1002/2016jb013057.

Poland, M. P., and D. Carbone (2018), Continuous gravity and tilt reveal anomalous pressure and density changes associated with gas pistoning within the summit lava lake of Kīlauea Volcano, Hawai‘i, *Geophysical Research Letters*, *45*(5), 2319–2327, doi:10.1002/2017GL076936.

Poland, M. P., and E. de Zeeuw-van Dalfsen (2019), Assessing seasonal changes in microgravity at Yellowstone caldera, *Journal of Geophysical Research: Solid Earth*, *124*(4), 4174–4188, doi:10.1029/2018JB017061.

Poland, M. P., A. Miklius, A. Jeff Sutton, and C. R. Thornber (2012), A mantle-driven surge in magma supply to Kīlauea Volcano during 2003 – 2007, *Nature Geoscience*, *5*(4), 295–300, doi:10.1038/ngeo1426.

Poland, M. P., A. Miklius, and E. K. Montgomery-Brown (2014), Magma supply, storage, and transport at shield-stage Hawaiian volcanoes, pp. 179–234, doi:10.3133/pp18015.

Poland, M. P., E. de Zeeuw-van Dalfsen, M. Bagnardi, and I. A. Johanson (2019), Post-collapse gravity increase at the summit of Kīlauea Volcano, Hawai‘i, *Geophysical Research Letters*, *46*(24), 14,430–14,439, doi:10.1029/2019GL084901.

Poland, M. P., D. Carbone, and M. R. Patrick (2021a), Onset and evolution of Kīlauea's 2018 flank eruption and summit collapse from continuous gravity, *Earth and Planetary Science Letters*, *567*, 117,003, doi:10.1016/j.epsl.2021.117003.

Poland, M. P., A. Miklius, I. A. Johanson, and K. R. Anderson (2021b), A decade of geodetic change at Kīlauea's summit – observations, interpretations, and unanswered questions from studies of the 2008 – 2018 Halema‘uma‘u eruption, *US Geological Survey Professional Paper, chap G*, doi:10.3133/pp1867g.

Poland, M. P., S. Hurwitz, J. P. Kauahikaua, E. K. Montgomery-Brown, K. R. Anderson, I. A. Johanson, M. R. Patrick, and C. A. Neal (2022), Rainfall an unlikely factor in Kīlauea's 2018 rift eruption, *Nature*, *602*(7895), E7–E10, doi:10.1038/s41586-021-04163-1.

Powell, M. J. (1964), An efficient method for finding the minimum of a function of several variables without calculating derivatives, *The Computer Journal*, *7*(2), 155–162, doi:10.1093/comjnl/7.2.155.

Ranta, E., S. A. Halldórsson, P. H. Barry, S. Ono, J. G. Robin, B. I. Kleine, A. Ricci, J. Fiebig, Á. E. Sveinbjörnsdóttir, and A. Stefánsson (2023), Deep magma degassing and volatile fluxes through volcanic hydrothermal systems: Insights from the Askja and Kverkfjöll volcanoes, Iceland, *Journal of Volcanology and Geothermal Research*, *436*, 107,776, doi:10.1016/j.jvolgeores.2023.107776.

Raspberry Shake, S.A. (2016), Raspberry Shake, doi:10.7914/SN/AM.

Rau, G. C. (2018), Hydrogeoscience/Pygtide: Pygtide V0.2, doi:10.5281/ZENODO.1346664.

Reilly, W. (1970), Adjustment of gravity meter observations, *New Zealand Journal of Geology and Geophysics*, *13*(3), 697–702, doi:10.1080/00288306.1970.10431341.

Reudink, R., R. Klees, O. Francis, J. Kusche, R. Schlesinger, A. Shabanloui, N. Sneeuw, and L. Timmen (2014), High tilt susceptibility of the Scintrex CG-5 relative gravimeters, *Journal of Geodesy*, *88*(6), 617–622, doi:10.1007/s00190-014-0705-0.

Ringler, A. T., M. Hagerty, J. Holland, A. Gonzales, L. S. Gee, J. Edwards, D. Wilson, and A. Baker (2015), The data quality analyzer: A quality control program for seismic data, *Computers & Geosciences*, *76*, 96–111, doi:10.1016/j.cageo.2014.12.006.

Rivalta, E., and P. Segall (2008), Magma compressibility and the missing source for some dike intrusions, *Geophysical Research Letters*, *35*(4), doi:10.1029/2007GL032521.

Rosat, S., and J. Hinderer (2018), Limits of detection of gravimetric signals on Earth, *Scientific Reports*, *8*(1), 1–8, doi:10.1038/s41598-018-33717-z.

Rosat, S., J. Hinderer, D. Crossley, and J. Boy (2004), Performance of superconducting gravimeters from long-period seismology to tides, *Journal of Geodynamics*, *38*(3-5), 461–476, doi:10.1016/j.jog.2004.07.005.

Ruigrok, E., J. Domingo-Ballesta, G.-J. van den Hazel, B. Dost, and L. Evers (2019), Groningen explosion database, *First Break*, *37*(8), 37–41, doi:10.3997/1365-2397.n0053.

Rymer, H. (1994), Microgravity change as a precursor to volcanic activity, *Journal of Volcanology and Geothermal Research*, *61*(3-4), 311–328, doi:10.1016/0377-0273(94)90011-6.

Rymer, H., and G. Brown (1989), Gravity changes as a precursor to volcanic eruption at Poas volcano, Costa Rica, *Nature*, *342*(6252), 902–905, doi:10.1038/342902a0.

Rymer, H., and E. Tryggvason (1993), Gravity and elevation changes at Askja, Iceland, *Bulletin of Volcanology*, *55*(5), 362–371, doi:10.1007/BF00301147.

Rymer, H., and G. Williams-Jones (2000), Volcanic eruption prediction: Magma chamber physics from gravity and deformation measurements, *Geophysical Research Letters*, *27*(16), 2389–2392, doi:10.1029/1999GL011293.

Rymer, H., J. Cassidy, C. A. Locke, and F. Sigmundsson (1998), Post-eruptive gravity changes from 1990 to 1996 at Krafla volcano, Iceland, *Journal of Volcanology and Geothermal Research*, *87*(1-4), 141–149, doi:10.1016/S0377-0273(98)00092-4.

Rymer, H., C. Locke, B. G. Ófeigsson, P. Einarsson, and E. Sturkell (2010), New mass increase beneath Askja volcano, Iceland–a precursor to renewed activity?, *Terra Nova*, *22*(4), 309–313, doi:10.1111/j.1365-3121.2010.00948.x.

Saibi, H., J. Gottsmann, and S. Ehara (2010), Post-eruptive gravity changes from 1999 to 2004 at Unzen volcano (Japan): A window into shallow aquifer and hydrothermal dynamics, *Journal of Volcanology and Geothermal Research*, *191*(1-2), 137–147, doi:10.1016/j.jvolgeores.2010.01.007.

Schippkus, S., M. Garden, and G. Bokelmann (2020), Characteristics of the Ambient Seismic Field on a Large-N Seismic Array in the Vienna Basin, *Seismological Research Letters*, *91*(5), 2803–2816, doi:10.1785/0220200153.

Scintrex Limited, C. (2012), CG-5 Scintrex autograv system operation manual.

Shahar Shani-Kadmiel, Frank Linde, Läslo Evers, and Bjorn Vink (2020), Einstein telescope seismic campaigns, doi:10.7914/SN/3T_2020.

Sigvaldason, G. E. (1979), Rifting, magmatic activity and interaction between acid and basic liquids, *Nordic Volcanological Institute*, *79*(03).

Sleeman, R. (2006), Three-channel correlation analysis: A new technique to measure instrumental noise of digitizers and seismic sensors, *Bulletin of the Seismological Society of America*, *96*(1), 258–271, doi:10.1785/0120050032.

Sleeman, R., H. Haak, M. Bos, and J. Van Gend (2000), Tidal tilt observations in the Netherlands using shallow borehole tiltmeters, *Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy*, *25*(4), 415–420, doi:10.1016/S1464-1895(00)00065-X.

Sparks, R., J. Biggs, and J. Neuberg (2012), Monitoring volcanoes, *Science*, *335*(6074), 1310–1311, doi:10.1126/science.1219485.

Sparks, R. S. J., L. Wilson, and H. Sigurdsson (1981), The pyroclastic deposits of the 1875 eruption of Askja, Iceland, *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *299*(1447), 241–273, doi:10.1098/rsta.1981.0023.

Spetzler, J., and B. Dost (2017), Hypocentre estimation of induced earthquakes in Groningen, *Geophysical Journal International*, *209*(1), 453–465, doi:10.1093/gji/ggx020.

Steim, J. (2015), Theory and observations – Instrumentation for global and regional seismology.

Stonebraker, M., P. Brown, D. Zhang, and J. Becla (2013), SciDB: A database management system for applications with complex analytics, *Computing in Science & Engineering*, *15*(3), 54–62, doi:10.1109/mcse.2013.19.

Strollo, A., P. Evans, A. Heinloo, S. Hemmleb, L. Hillmann, K.-H. Jäckel, J. Quinteros, J. Saul, R. Zaccarelli, T. Zieke, et al. (2020), GEOFON annual report 2019, doi:10.2312/GFZ.B103-20035.

Sturkell, E., and F. Sigmundsson (2000), Continuous deflation of the Askja caldera, Iceland, during the 1983–1998 noneruptive period, *Journal of Geophysical Research: Solid Earth*, *105*(B11), 25,671–25,684, doi:10.1029/2000JB900178.

Sturkell, E., F. Sigmundsson, and P. Einarsson (2003), Recent unrest and magma movements at Eyjafjallajökull and Katla volcanoes, Iceland, *Journal of Geophysical Research: Solid Earth*, *108*(B8), doi:10.1029/2001JB000917.

Sturkell, E., F. Sigmundsson, and R. Slunga (2006), 1983–2003 decaying rate of deflation at Askja caldera: pressure decrease in an extensive magma plumbing system at a spreading plate boundary, *Bulletin of Volcanology*, *68*(7), 727–735, doi:10.1007/s00445-005-0046-1.

Sugihara, M., and T. Ishido (2008), Geothermal reservoir monitoring with a combination of absolute and relative gravimetry, *Geophysics*, *73*(6), WA37–WA47, doi:10.1190/1.2991105.

Sutton, A., and T. Elias (2014), One hundred volatile years of volcanic gas studies at the Hawaiian Volcano Observatory, *US Geological Survey Professional Paper*, doi:10.3133/pp18017.

Swanson, D. A. (1972), Magma supply rate at Kīlauea Volcano, 1952 – 1971, *Science*, *175*(4018), 169–170, doi:10.1126/science.175.4018.169.

Takasu, T. (2013), RTKLIB ver. 2.4. 2 manual, *RTKLIB: An Open Source Program Package for GNSS Positioning*, pp. 29–49.

Thorarinsson, S., and G. E. Sigvaldason (1962), The eruption in Askja, 1961; a preliminary report, *American Journal of Science*, *260*(9), 641–651, doi:10.2475/ajs.260.9.641.

Trani, L., M. Koymans, M. Atkinson, R. Sleeman, and R. Filgueira (2017), WFCatalog: A catalogue for seismological waveform data, *Computers & Geosciences*, *106*, 101–108, doi:10.1016/j.cageo.2017.06.008.

Tryggvason, E. (1989), Ground deformation in Askja, Iceland: its source and possible relation to flow of the mantle plume, *Journal of Volcanology and Geothermal Research*, *39*(1), 61–71, doi:10.1016/0377-0273(89)90021-8.

Uieda, L., D. Tian, W. Leong, L. Toney, W. Schlitzer, M. Grund, D. Newton, M. Ziebarth, M. Jones, and P. Wessel (2021), PyGMT: A Python interface for the Generic Mapping Tools, doi:10.5281/zenodo.7772533.

USGS Hawaiian Volcano Observatory (HVO) (1956), Hawaiian Volcano Observatory Network, doi:10.7914/SN/HV.

Vajda, P., I. Foroughi, P. Vaníček, R. Kingdon, M. Santos, M. Sheng, and M. Goli (2020), Topographic gravimetric effects in earth sciences: Review of origin, significance and implications, *Earth-Science Reviews*, *211*, 103,428, doi:10.1016/j.earscirev.2020.103428.

Vajda, P., P. Zahorec, C. A. Miller, H. Le Mével, J. Papčo, and A. G. Camacho (2021), Novel treatment of the deformation–induced topographic effect for interpretation of spatiotemporal gravity changes: Laguna del Maule (Chile), *Journal of Volcanology and Geothermal Research*, *414*, 107,230, doi:10.1016/j.jvolgeores.2021.107230.

Van Camp, M., and P. Vauterin (2005), TSoft: graphical and interactive software for the analysis of time series and Earth tides, *Computers & Geosciences*, *31*(5), 631–640, doi:10.1016/j.cageo.2004.11.015.

Van Camp, M., O. de Viron, A. Watlet, B. Meurers, O. Francis, and C. Caudron (2017), Geophysics from terrestrial time-variable gravity measurements, *Reviews of Geophysics*, *55*(4), 938–992, doi:10.1002/2017RG000566.

Van Der Walt, S., S. C. Colbert, and G. Varoquaux (2011), The NumPy array: a structure for efficient numerical computation, *Computing in Science & Engineering*, *13*(2), 22, doi:10.1109/MCSE.2011.37.

Van Rossum, G., and F. L. Drake (2009), *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA.

Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors (2020), SciPy 1.0: Fundamental algorithms for scientific computing in Python, *Nature Methods*, *17*, 261–272, doi:10.1038/s41592-019-0686-2.

Wang, R., S. Parolai, M. Ge, M. Jin, T. R. Walter, and J. Zschau (2013), The 2011 Mw 9.0 Tohoku earthquake: Comparison of GPS and strong-motion data, *Bulletin of the Seismological Society of America*, *103*(2B), 1336–1347, doi:10.1785/0120110264.

Wang, T., Y. Zheng, F. Pulvirenti, and P. Segall (2021), Post-2018 caldera collapse re-inflation uniquely constrains Kīlauea's magmatic system, *Journal of Geophysical Research: Solid Earth*, *126*(6), doi:10.1029/2021jb021803.

Warburton, R., H. Pillai, and R. Reineman (2010), Initial results with the new GWR iGrav superconducting gravity meter, in *Extended Abstract Presented at 2nd Asia Workshop on Superconducting Gravimetry Taipei, Taiwan.*

Wech, A. G., W. A. Thelen, and A. M. Thomas (2020), Deep long-period earthquakes generated by second boiling beneath Mauna Kea volcano, *Science*, *368*(6492), 775–779, doi:10.1126/science.aba4798.

Weiser, B., H. Pfutzner, and J. Anger (2000), Relevance of magnetostriction and forces for the generation of audible noise of transformer cores, *IEEE Transactions on Magnetics*, *36*(5), 3759–3777, doi:10.1109/20.908346.

Welch, P. (1967), The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, *IEEE Transactions on Audio and Electroacoustics*, *15*(2), 70–73, doi:10.1109/tau.1967.1161901.

Wenzel, H.-G. (1996), The nanogal software: Earth tide data processing package ETERNA 3.30, *Bull. Inf. Marées Terrestres*, *124*, 9425–9439.

Werner, C., U. Wegmüller, T. Strozzi, and A. Wiesmann (2000), Gamma SAR and interferometric processing software, in *Proceedings of the ERS-ENVISAT Symposium, Gothenburg, Sweden*, vol. 1620, p. 1620, Citeseer.

Wessel, P., J. F. Luis, L. Uieda, R. Scharroo, F. Wobbe, W. H. F. Smith, and D. Tian (2019a), The Generic Mapping Tools version 6, doi:10.5281/zenodo.3407866, Funded by US National Science Foundation grants OCE-1558403 and EAR-1829371.

Wessel, P., J. Luis, L. Uieda, R. Scharroo, F. Wobbe, W. H. Smith, and D. Tian (2019b), The Generic Mapping Tools version 6, *Geochemistry, Geophysics, Geosystems*, *20*(11), 5556–5564, doi:10.1029/2019GC008515.

Wieser, P. E., M. Edmonds, J. Maclennan, and J. Wheeler (2020), Microstructural constraints on magmatic mushes under Kīlauea Volcano, Hawaii, *Nature Communications*, *11*(1), 1–14, doi:10.1038/s41467-019-13635-y.

Wijaya, D. D., N. A. Muhamad, K. Prijatna, V. Sadarviana, D. A. Sarsito, A. Pahlevi, E. D. Variandy, and W. Putra (2019), pyGABEUR-ITB: A free software for adjustment of relative gravimeter data, *Geomatika*, *25*(2), 95–102, doi:10.24895/jig.2019.25-2.991.

Winder, T. (2022), Tectonic earthquake swarms in the Northern Volcanic Zone, Iceland, doi:10.17863/CAM.82505.

Wolin, E., and D. E. McNamara (2019), Establishing high-frequency noise baselines to 100Hz based on millions of power spectra from IRIS MUSTANG, *Bulletin of the Seismological Society of America*, *110*(1), 270–278, doi:10.1785/0120190123.

Xia, J., and R. D. Miller (2000), Design of a hum filter for suppressing power-line noise in seismic data, *Journal of Environmental & Engineering Geophysics*, *5*(2), 31–38, doi:10.4133/JEEG5.2.31.

Xu, P., C. Shi, R. Fang, J. Liu, X. Niu, Q. Zhang, and T. Yanagidani (2013), High-rate precise point positioning (PPP) to measure seismic wave motions: an experimental comparison of GPS PPP with inertial measurement units, *Journal of Geodesy*, *87*(4), 361–372, doi:10.1007/s00190-012-0606-z.

Yun, S., P. Segall, and H. Zebker (2006), Constraints on magma chamber geometry at Sierra Negra Volcano, Galápagos Islands, based on InSAR observations, *Journal of Volcanology and Geothermal Research*, *150*(1-3), 232–243, doi:10.1016/j.jvolgeores.2005.07.009.

Zurek, J., G. William-Jones, D. Johnson, and A. Eggers (2012), Constraining volcanic inflation at Three Sisters Volcanic Field in Oregon, USA, through microgravity and deformation modeling, *Geochemistry, Geophysics, Geosystems*, *13*(10), Q10,013, doi:10.1029/2012gc004341.

List of symbols and abbreviations

Symbols

Boldface type indicates vector quantities.

Symbol	Description	Units
ω	Angular Frequency	rad/s
h	Damping factor	-
$\mathbf{r}$	Euclidean distance	m
f	Frequency	s^{-1}
A	Full-load amplitude	-
G_e	Generator constant	$V\,m^{-1}\,s^{-1}$
M	Mass	kg
δm	Mass (change)	kg
N	Number of samples	-
n	Number of effective bits	-
PSD	Power Spectral Density (acceleration)	$m^2\,s^{-4}\,Hz^{-1}$
η	Proxybits	-
Δ	Quantization interval	-
ϵ_{rms}	Quantization noise	$m\,s^{-2}$
f_s	Sampling frequency	s^{-1}
T	Sampling interval	s
G	Universal gravitational constant	$m^3\,kg^{-1}\,s^{-2}$
g	Vert. component of gravitational acceleration	$m\,s^{-2}$
δg	Vert. component of gravitational acceleration (change)	$m\,s^{-2}$

Abbreviations

ADC	Analog-to-Digital Converter
API	Application Programming Interface
BCFAG	Bouguer corrected free-air gradient
BGI	International Gravimetric Bureau
CTBTO	Comprehensive Nuclear-Test-Ban Treaty
DFT	Discrete Fourier Transform
EIDA	European Integrated Data Archive
ENF	Electrical Network Frequency
FAG	Free-air gradient
FAIR	Findability, Accessibility, Interoperability, and Reuse
FFT	Fast Fourier Transform
GMT	Generic Mapping Tools
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HMMR	Halemaʻumaʻu reservoir
HVO	Hawaiian Volcano Observatory
IDC	International Data Centre
IERS	International Earth Rotation and Reference Systems Service
InSAR	Interferometric synthetic-aperture radar
KNMI	Royal Netherlands Meteorological Institute
NHNM	New High Noise Model
NLNM	New Low Noise Model
NSAN	Netherlands Seismic and Acoustic Network
NTP	Network Time Protocol
ORFEUS	Observatories & Research Facilities for European Seismology
PPP	Precise Point Position
PPSD	Probabilistic Power Spectral Density
PSD	Power Spectral Density
SCR	South Caldera reservoir
SG	Superconducting Gravimeter
USGS	United States Geological Survey
UTC	Universal Time Coordinated